



HAL
open science

Musical and verbal auditory cognition : child development and neurophysiological markers

Jérémie Ginzburg

► **To cite this version:**

Jérémie Ginzburg. Musical and verbal auditory cognition : child development and neurophysiological markers. Neuroscience. Université Claude Bernard - Lyon I, 2023. English. NNT : 2023LYO10215 . tel-04609221

HAL Id: tel-04609221

<https://theses.hal.science/tel-04609221v1>

Submitted on 12 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de Doctorat de L'Université Claude Bernard Lyon 1

Ecole Doctorale N° 476
Neurosciences et Cognition
Discipline : Neurosciences

Soutenue publiquement le 16/11/2023, par :
Jérémy Ginzburg

Musical and verbal auditory cognition: child development and neurophysiological markers

Devant le jury composé de :

Dre Marie Gomot, CR INSERM Tours (rapporteure)

Dr Daniel Pressnitzer, DR CNRS Paris (rapporteur)

Dr Pascal Barone, DR CNRS Toulouse (examinateur)

Dre Gaën Plancher, MC Université Lyon 2 (examinatrice)

Pr Eric Truy, PU-PH Université Lyon 1 (président)

Dre Anne Caclin, CR INSERM Lyon (directrice de thèse)

« Plus un corps tombe moins vite, moins sa vitesse est plus grande. »

Pierre Ginzburg - *Une certaine vision de la science*

« La musique c'est le mouvement invisible. »

Nancy Huston – *Ligne de faille*

« Il semble qu'il existe dans le cerveau une zone tout à fait spécifique qu'on pourrait appeler la mémoire poétique et qui enregistre ce qui nous a charmés, ce qui nous a émus, ce qui donne à notre vie sa beauté. »

Milan Kundera – *L'insoutenable légèreté de l'être*

Remerciements

Anne (Caclin, il faut préciser avec tous ces homonymes), il est difficile d'exprimer en peu de mots tout ce que tu m'as apporté ces dernières années. Je suis fier d'avoir parcouru cette thèse avec toi, fier d'avoir été ton premier encadrement « en solo », fier d'avoir drainé ton cerveau de toutes les connaissances dont j'avais besoin. J'aimerais que tous les thésard.e.s puissent être formé par une personne aussi bienveillante, exigeante et patiente que toi. J'admire ta capacité à toujours être disponible pour tes étudiant.e.s, efficace dans ton travail, intransigeante du point de vue scientifique, empathique dans tes rapports professionnels, tout en gardant une vie personnelle équilibrée. En une phrase : quand je serai grand je serai comme toi.

Je tiens à exprimer ma gratitude à Marie Gomot et Daniel Pressnitzer qui ont accepté de relire et évaluer cette thèse ainsi que les membres du jury Pascal Barone, Gaën Plancher et Eric Truy.

Merci à toutes les collaboratrices des différentes études présentées dans ce projet : Barbara, Annie, NaN, Laura. Merci également à Gaën qui m'a permis de découvrir l'enseignement à la fac, expérience que je renouvellerai avec plaisir.

Je tiens à remercier toutes les « mamans » de l'équipe PAM qui ont réussi à faire de cette équipe de recherche un espace de respect, de dialogue et de qualité scientifique qui me semble rare et digne d'être mentionné. Merci à Perrine sans qui, si tu n'étais pas revenu me chercher par la peau du cou pour faire une thèse, je n'aurais pas accompli ce travail, merci pour notre amitié et nos collaborations fluctuantes mais toujours passionnantes. Merci à Lesly (peut-être pas une des mamans mais plutôt une grande sœur), Aurélie (ma première directrice de thèse officielle et bien plus), Karine (tu me dois toujours un apéro), Annie (merci pour mon premier congrès international et toutes nos

Remerciements

collaborations), NaN (et sa présence magique qui fait fonctionner les ordinateurs), Laure (à l'origine de ma culture polysomnographique).

Merci à la première génération de doctorant.e.s PimPamPoum. Je garde notre conversation messenger près de mon coeur, si importante pendant les confinements, couvre-feu, réunions, et lieu de narration des histoires de douche, ainsi que tous les autres moments qu'on a passé ensemble : Roxane, Lou, Salomé, Francesca, Rémy, Agathe. Merci à Aurélien et Caliani, la deuxième génération de doctorant.e.s avec qui ça a été un immense plaisir de rigoler, d'apprendre, d'aller en congrès et de souffrir (mais pas que).

Faisant partie des meubles de DYCOG, j'ai énormément de gens à remercier pour avoir rendu ces années exceptionnellement riches, il me faudrait beaucoup plus de pages pour rendre à chacun.e ce que vous m'avez apporté, ne vous vexez pas si ces mots de reflètent pas la qualité de ce qu'on a vécu ensemble: Flo, Maxime, Benoît, Alex, Nicole, Romain B, Manu, Nicolas, Gaëtan, Stefano, Oussama, Arnaud, Romain Q, Anne K, Judith, Mélodie, Léa, Pauline, Sébastien, Denis, Julie, Pierre, Côme, Sotiris, Letizia, Dafina, Benjamin, Maryne. Un remerciement spécial pour Martine et Fatmah qui m'ont aidé un nombre incalculable de fois et avec qui j'ai pris tant de plaisir à discuter et apprendre sur des sujets non-scientifiques, mais tout aussi important.

Merci à tous les stagiaires que j'ai encadré pour tous les cafés que vous auriez dû m'apporter si j'en buvais : Eva, Elise, RomRom, Lilou, Jeannine, Laurine, Solène, Thibault, Florian.

Merci à Philippe et Roxane de m'avoir accueilli à Québec pendant 4 mois, c'était tellement l'fun, j'arrive bientôt ! Merci à tous les ami.e.s sur place qui ont rendu ce séjour si mémorable : Felix, Anna, Adrian, Renata, Florian.

Remerciements

Merci à Robert et Philippe de me permettre de continuer cette aventure scientifique chez nos cousins Québécois.

Merci à Milena, Raphaël et Doudou, les piliers de ma vie affective. Merci Maman d'avoir permis l'enclenchement de la rédaction de cette thèse (et de ma vie, et de tout ce qu'il y a entre les deux). Merci à Nanon, que j'espère rendre fier avec ce travail. Merci à toute ma famille et équivalents : Rémi, Cyril, Fabienne, Julien, Clara, Naïs, Emilie, Loïc, Mimi, Serge, Pascal. Un énorme bisou à Adèle (et à l'échantillon en cours de culture), Zoé et Lola qui, soit ont déjà passé mes expériences, soit y passeront un de ces jours. Merci à Isam, Kéké et Chastou, pour votre amitié inconditionnelle pendant les années passées et à venir.

Un paragraphe de remerciement spécial à Andrea, qui a accepté de faire pencher la balance du patriarcat du mauvais côté pendant ces quelques mois de rédaction. Merci pour ta présence quotidienne (since the beniging), ton soutien, tes conseils, tes relectures, ta compréhension, ta douceur et bien d'autres choses. J'ai hâte de passer à la prochaine étape Montréalaise avec toi (et Pois Chiche). Je tiens aussi à remercier toute la famille Diaz-Barriga Yanez de m'accueillir aussi facilement dans votre famille et de me soutenir de loin.

Enfin, merci à Papa, dont je regrette tant l'absence pour cette dernière étape.

Abstract

Central auditory processes play a pivotal role in various aspects of everyday life and are particularly important for the comprehensive development of children, as they form the foundation for effective communication, language, and reading skills. Children with language-related disorders indeed often show impairments in auditory cognitive processes such as speech-in-noise perception, auditory attention, and auditory short-term memory (STM). This fact stresses the need to screen central auditory processes early in childhood in order to provide information about potential future language and reading acquisition difficulties. In particular, numerous studies have suggested that auditory STM plays a substantial part in the core phonological deficit that language-related learning disorders are thought to arise from. Moreover, the consistently observed auditory STM impairments in learning disorders does not only concern STM for verbal material but also musical material, suggesting for some cases a domain-general impairment of STM that goes beyond phonological material. While behavioral studies during development are important to increase our knowledge of the maturation steps that such auditory processes undergo, the neurophysiological investigation of these processes allow for a more in-depth and precise mapping of their dynamics during development. One relatively new neuroimaging technique, functional near-infrared spectroscopy (fNIRS), stands out for such investigations as it is silent and has minimal physical constraints, making it an ideal candidate for the investigation of the auditory modality in children and clinical populations. In the present work, we first conducted two behavioral studies in which we tested 100 children from 5 to 10 years old on speech-in-noise perception and auditory STM for musical and verbal material. We showed that both processes were not mature by 10 years of age and that both processes undergo a decisive maturation step around 7 years of age.

Abstract

Moreover, we uncovered evidence for shared (e.g., similar developmental trajectory) and distinct (e.g., correlation between musical STM and speech-in-noise perception, but not verbal STM) mechanisms for musical and verbal STM. We then conducted two neurophysiological studies in which we investigated auditory STM for musical and verbal material using fNIRS. In the first study, we conducted two experiments on healthy adults in which we showed the involvement of lateral prefrontal regions in auditory STM tasks when compared to perception tasks, thus confirming previous neuroimaging literature. Moreover, we showed that the inferior frontal gyrus (IFG) parametrically responded to the increase of memory load for musical and verbal STM. Finally, we uncovered a different parametric involvement of the IFG for musical and verbal material, suggesting that fNIRS can reliably capture dynamic changes in relation to the strategy used for musical and verbal material. In the second ongoing fNIRS study, we are investigating the lateral prefrontal engagement in auditory STM for musical and verbal material in typically developing children (TD) and children with language-related learning disorders. Initial evidence in 11 TD children confirms the interest of fNIRS to study the involvement of prefrontal areas in auditory STM during typical and atypical development. Overall, this PhD work brings knowledge about the development of central auditory processes, highlighting their importance in children's harmonious language and reading acquisition. Moreover, we provide evidence for the well-suitedness of fNIRS to explore central auditory processes in healthy adults and children with typical and atypical development.

Résumé

Les processus auditifs centraux jouent un rôle essentiel dans de multiples aspects de la vie quotidienne et sont particulièrement importants pour le développement de l'enfant, car ils constituent la base d'une communication efficace et des compétences en lecture. Les enfants souffrant de troubles du langage présentent souvent des déficits de cognition auditive, notamment pour la perception de la parole dans le bruit, l'attention auditive et la mémoire à court terme (MCT) auditive. Ceci souligne la nécessité d'examiner les processus auditifs centraux tôt dans l'enfance afin d'alerter sur les futures potentielles difficultés d'acquisition du langage et de la lecture. En particulier, de nombreuses études ont suggéré que la MCT auditive joue un rôle important dans le déficit phonologique qui serait à l'origine des troubles de l'apprentissage associés au langage. En outre, les déficits de la MCT auditive régulièrement associés aux troubles de l'apprentissage ne concernent pas seulement la MCT pour le matériel verbal, mais aussi pour le matériel musical, ce qui suggère, dans certains cas, un déficit de MCT plus général qui ne concerne pas que le matériel phonologique. Bien que les études comportementales au cours du développement soient importantes pour améliorer notre connaissance des étapes de maturation que subissent ces processus auditifs, l'étude neurophysiologique de ces processus permet une cartographie plus approfondie et plus précise de leur dynamique au cours du développement. Une technique de neuro-imagerie relativement nouvelle, la spectroscopie fonctionnelle proche-infrarouge (fNIRS), est particulièrement adaptée pour de telles investigations car sa nature silencieuse et le peu de contraintes associées en font un candidat idéal pour l'investigation de la modalité auditive chez les enfants et les populations cliniques. Dans ce travail de thèse, nous avons d'abord mené deux études comportementales dans lesquelles nous avons testé 100 enfants âgés de 5 à 10

Résumé

ans sur leur capacité de perception de la parole dans le bruit et sur leur MCT auditive pour du matériel musical et verbal. Nous avons montré que ces deux processus n'étaient pas matures à l'âge de 10 ans et qu'ils subissaient une étape de maturation décisive vers l'âge de 7 ans. De plus, nous avons fourni des preuves de l'existence de mécanismes communs (e.g., une trajectoire développementale similaire) et distincts (e.g., une corrélation entre les performances de perception de la parole dans le bruit et de MCT musicale, mais pas avec la MCT verbale) pour les processus sous-tendant la MCT musicale et verbale. Nous avons ensuite mené deux études neurophysiologiques dans lesquelles nous avons examiné la MCT auditive pour le matériel musical et verbal à l'aide de la fNIRS. Dans la première étude, nous avons mené deux expériences chez des adultes sains dans lesquelles nous avons montré l'engagement des régions préfrontales latérales dans des tâches de MCT auditive par rapport à des tâches de perception, confirmant ainsi des résultats d'études de neuro-imagerie antérieures. En outre, nous avons montré que le gyrus frontal inférieur (IFG) répondait de manière paramétrique à l'augmentation de la charge mnésique pour les tâches de MCT musicales et verbales. Enfin, nous avons observé une implication paramétrique différente de l'IFG pour le matériel musical et verbal, suggérant que la fNIRS peut capturer de manière fiable les changements dynamiques en relation avec les stratégies utilisées pour le matériel musical et verbal. Dans la deuxième étude, en cours de réalisation, nous étudions l'engagement des régions préfrontales latérales dans la MCT auditive pour du matériel musical et verbal chez des enfants au développement typique et des enfants souffrant de troubles de l'apprentissage liés au langage. Les premiers résultats obtenus chez 11 enfants au développement typique confirment l'intérêt de la fNIRS pour l'étude de l'implication des aires préfrontales dans la MCT auditive au cours du développement typique et

Résumé

atypique. Dans l'ensemble, ce travail de thèse a permis d'acquérir des connaissances sur le développement des processus auditifs centraux au cours de l'enfance, soulignant leur importance dans l'acquisition harmonieuse du langage et de la lecture par les enfants. De plus, nous avons démontré que la fNIRS est bien adaptée à l'exploration des processus auditifs centraux chez les adultes et chez les enfants au développement typique et atypique.

Contents

Remerciements	3
Abstract	6
Résumé.....	8
Contents.....	11
Introduction	15
1 Theoretical background.....	19
1.1 Development of auditory cognition.....	19
1.1.1 Auditory scene analysis (ASA)	19
<i>1.1.1.1 ASA: mechanisms.....</i>	<i>19</i>
<i>1.1.1.2 Speech recognition in speech, a special case of ASA</i>	<i>21</i>
<i>1.1.1.3 Development of speech-in-noise perception.....</i>	<i>25</i>
1.1.2 Auditory attention.....	30
<i>1.1.2.1 Auditory selective attention</i>	<i>30</i>
<i>1.1.2.2 Development of auditory selective attention.....</i>	<i>32</i>
1.1.3 Auditory short-term memory	35
<i>1.1.3.1 Memory.....</i>	<i>35</i>
<i>1.1.3.2 Working memory</i>	<i>39</i>
<i>1.1.3.3 Short-term memory (STM).....</i>	<i>43</i>
<i>1.1.3.4 Auditory STM for verbal and musical material.....</i>	<i>47</i>
<i>1.1.3.5 Development of auditory STM.....</i>	<i>55</i>
1.2 Objective markers of auditory cognition: functional near-infrared spectroscopy.....	62
1.2.1 Auditory cognition: cerebral networks	63
<i>1.2.1.1 Cortical networks of auditory processing.....</i>	<i>63</i>
<i>1.2.1.2 Cerebral correlates of speech processing in adverse conditions</i>	<i>68</i>
1.2.2 Auditory STM: cerebral networks.....	71
<i>1.2.2.1 Cerebral correlates of verbal STM.....</i>	<i>71</i>
<i>1.2.2.2 Cerebral correlates of musical STM.....</i>	<i>74</i>
<i>1.2.2.3 Comparison between verbal and musical STM.....</i>	<i>75</i>
1.2.3 Functional Near-Infrared Spectroscopy (fNIRS)	79

Contents

1.2.3.1	<i>History and instrumentation of fNIRS</i>	79
1.2.3.2	<i>Physical principles</i>	82
1.2.3.3	<i>Experimental design: constraints from hemodynamics</i>	85
1.2.3.4	<i>Pre-processing of fNIRS data</i>	87
1.2.3.5	<i>Data analysis</i>	92
1.2.3.6	<i>Limitations of fNIRS</i>	94
1.2.4	fNIRS in developmental science	95
1.2.5	Exploring auditory STM with fNIRS	98
	Objectives.....	102
2	Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception	105
2.1	General Introduction	105
2.2	Article	105
	Abstract.....	106
	Methods	115
	Results	124
	Discussion	133
	References	142
	Supplementary material	149
3	Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory	151
3.1	General introduction	151
3.2	Article.....	151
	Research Highlights.....	153
	Abstract.....	153
	Introduction	155
	Methods	160
	Results	172
	Discussion	181
	Conclusion	188
	References	189
	Supplementary materials.....	202

Contents

4	Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study.....	207
4.1	General introduction	207
4.2	Article.....	208
	Abstract.....	210
	Introduction.....	212
	Experiment 1	220
	<i>Methods</i>	220
	<i>Results</i>	230
	<i>Discussion</i>	233
	Experiment 2	235
	<i>Methods</i>	235
	<i>Results</i>	239
	<i>Discussion</i>	245
	General Discussion	247
	References	254
	Supplementary material	262
5	Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.	270
5.1	General introduction	270
5.2	Report on preliminary results.....	270
	Introduction.....	270
	Methods	282
	Results	292
	Discussion	298
	References	301
	Supplementary material	309
6	General Discussion	311
6.1	Development of auditory cognition.....	311
6.2	Musical and verbal STM	314
6.3	The fNIRS method: interests and limitations	317
6.4	Perspectives for learning disorders.....	322

Contents

Conclusions	326
References	327

Introduction

Auditory processing involves a sequence of analysis stages, starting with peripheral auditory mechanisms responsible for the creation of a first neuronal representation of sounds, and moving towards more intricate phases where sound processing leads to the perception, retention, and identification of acoustic objects (Litovsky, 2015). In recent years, there has been a growing interest on the cognitive processes that underlie auditory processing at the central level (i.e., auditory cognition), including their pivotal role in development (Litovsky, 2015; Werner et al., 2012). Many children encounter unexpected challenges in understanding and producing spoken language despite normal intelligence, hearing ability, opportunities for language learning, and in the absence of neurological disorders. Researchers have long identified the processing of phonological material as a fundamental factor in language-related learning disorders such as dyslexia and developmental language disorders (DLD), suggesting that individuals with learning disorders struggle with the representation, storage, and retrieval of speech sounds (Bradley & Bryant, 1978; Ramus, 2003). Moreover, it has been suggested that these auditory processing difficulties extend beyond speech material (Couvignou et al., 2019, 2023; Couvignou & Kolinsky, 2021; Weiss et al., 2014; Ziegler et al., 2012).

Auditory scene analysis (ASA), including in particular speech-in-noise perception, auditory attention, and auditory short-term memory (STM) are some of the auditory cognition processes that have been demonstrated to be impaired in learning disorders. ASA, the ability to make perceptual sense of a complex sound environment, is a key component of auditory cognition that enables the segregation of sounds into perceptual streams and, in association with auditory selective attention, allows the priority processing of a sound

Introduction

source against irrelevant ones. Approximately 7% of children show specific difficulties in listening in noise despite normal audiograms (Wilson & Arnott, 2013) and this difficulty has been consistently observed in children with DLD (Ziegler et al., 2009, 2011). Other higher-order auditory functions have also been shown to be impaired in learning disorders. Auditory STM plays a crucial role in everyday life as it enables the active storage of relevant information that are, for example, central to follow a conversation, read and understand sentences, listen to music etc. Deficits in verbal STM have been consistently demonstrated as a hallmark of dyslexia (Majerus & Cowan, 2016) and DLD (Nithart et al., 2009). However, auditory STM impairments do not seem to be limited to verbal material as STM for musical tones has been shown to be impaired in dyslexic children (Forgeard et al., 2008; Ziegler et al., 2012). Furthermore, recent research has revealed a substantial comorbidity between dyslexia and congenital amusia (Couvignou et al., 2023), in keeping with a general deficit in auditory STM in (at least some cases of) dyslexia rather than one limited to speech sounds. These findings stress the need to extend our knowledge about the development of auditory cognition processes to allow for better screening, prevention, and rehabilitation of auditory cognition deficits notably in neurodevelopmental disorders.

A complementary approach to the use of behavioral protocols for understanding auditory cognition in children and adults is the investigation of its objective markers using electrophysiological and neuroimaging techniques that allow the exploration of the ongoing brain physiological dynamics during specific auditory processes. Scalp electroencephalography (EEG) and magnetoencephalography (MEG) have been used to characterize the fast-changing temporal dynamics of auditory cognition by the exploration of auditory evoked potentials/fields while positron emission tomography (PET)

Introduction

and functional magnetic resonance imaging (fMRI) have been used to precisely locate brain structures associated with specific processes. However, all these techniques present challenges when investigating children and clinical populations. Most of the previously mentioned techniques are highly sensitive to a participant's movements, and in most of them, the experimental setup can be particularly distressing for children and clinical populations (Ferreri et al., 2014). One emerging technology, functional near-infrared spectroscopy (fNIRS), a technique measuring cortical hemodynamic changes in response to cognitive events, allows to overcome some of the aforementioned difficulties. It provides a less restrictive experimental environment and has a higher tolerance to movements and muscular artifacts as compared to any other techniques (Aslin & Mehler, 2005). Furthermore, the noise generated by fMRI can pose obstacles when investigating the auditory domain, especially for individuals facing challenges in language processing because it requires their engagement in a noisy environment. In contrast, fNIRS is fully silent during measurements (Butler et al., 2020; Hancock et al., 2023).

The objectives of this PhD were two-fold: (1) we conducted behavioral investigations into the developmental trajectory of two critical components of auditory cognition: speech-in-noise perception and auditory STM for musical and verbal material. These processes have been demonstrated to undergo important maturation steps immediately following the acquisition of reading skills (around 6-7 years of age) until adolescence. Given the advantageous access to school-aged children (i.e., within a single school environment), our studies were centered on participants aged 5 to 10 years; (2) we explored objective markers of auditory STM for musical and verbal material using the fairly new and well-suited fNIRS neuroimaging technique. Our focus on auditory STM stems from its pivotal role in language and reading acquisition, consistently

Introduction

exhibiting impairment in language-related learning disorders. As the fNIRS research field on auditory STM in adults and children is relatively limited to date, we carried out two separate studies. The first one with healthy adults aimed to confirm the suitability of fNIRS for the identification of objective markers of auditory STM, and the second one aimed to explore these objective markers in typically developing children and children with language-related learning disorders.

The next theoretical background section will articulate around the two previously presented objectives. In a first subsection dedicated to the developmental trajectory of auditory cognition (1.1), we will review current knowledge about speech-in-noise perception (1.1.1) and auditory STM (1.1.3) and their development during school years. In between these two sections, we will briefly review a key auditory process that plays a significant role in both processes, auditory selective attention (1.1.2). In the second subsection dedicated to the objective markers of auditory cognition (1.2), we will review current knowledge of the cerebral networks subtending auditory cognition and speech processing (1.2.1). We will then focus on the cortical networks subtending auditory STM for musical and verbal material (1.2.2). Next, we will provide a comprehensive overview of the fNIRS neuroimaging technique (1.2.3), examine its primary applications in developmental science (1.2.4), and finally delve into the use of fNIRS to investigate auditory STM (1.2.5).

1 Theoretical background

1.1 Development of auditory cognition

The aim of this section is to present the current state of the art on three fundamental processes of auditory cognition and their development in childhood: auditory scene analysis (ASA), auditory attention, and auditory short-term memory (STM). For each of them, an overview of cognitive psychology models will be presented followed by a presentation of the current knowledge about their development throughout school years and the factors involved in their development.

1.1.1 Auditory scene analysis (ASA)

ASA is the cognitive process that underlies our ability to effortlessly segregate and interpret sounds from complex auditory environments (Bregman, 1994). This phenomenon allows the human brain to (1) parse incoming acoustic information into distinct perceptual streams and (2) identify these streams. This enables us to recognize individual sound sources, distinguish background noise, and understand multiple sound sources simultaneously (Bregman, 1994). The auditory system achieves this by employing various acoustic cues, such as spatial location, pitch, timbre, and temporal patterns, which aid in organizing and grouping auditory elements into meaningful and coherent representations.

1.1.1.1 *ASA: mechanisms*

Bergman (1994) proposed a definition of ASA as processes whereby listeners segregate incoming acoustic waveforms into multiple perceptual streams. This division hinges on two key mechanisms: a *primitive scene analysis* based on the segregation of incoming acoustic waveforms on the basis of general acoustic characteristics. It is usually accepted that segregating sounds using these basic mechanisms does not necessarily require prior listening experience.

Auditory scene analysis (ASA)

ASA also involves a subsequent *schema-based analysis*, which relies on learned schemas derived from a listener's previous experiences with sounds. Many studies on ASA focus on identifying and manipulating acoustic cues believed to be associated with basic grouping mechanisms.

The objective of ASA is often described as auditory stream segregation, which involves the ability to perceptually group incoming waveforms into distinct auditory streams on the basis of acoustic cues that foster coherence between elements across time (Litovsky, 2015). Over the past several decades, research has extensively explored various acoustic cues that facilitate the formation of auditory streams. These investigations have shown that adults present the ability to segregate incoming sounds based on cues such as spatial separation, frequency separation, spectral profile, talker sex, onset and offset times, temporal modulation, and harmonicity (Bregman, 1994). Notably, temporal onsets and offsets have emerged as particularly robust cues for auditory stream segregation. In other words, when frequency components start and stop simultaneously, they are more likely to originate from the same sound source, in contrast to components with different onset or offset times.

In real-life situations, auditory stream segregation often occurs in the context of segregating one relevant sound source from other irrelevant sound sources (Cherry, 1953). The detrimental effect of irrelevant sound sources that compete with a target source resulting in ambiguous perception or wrong recognition of the target source is known as “masking” (Kidd et al., 2008). The modern investigation of masking traces its roots back to Wegel and Lane (1924), who conducted a series of influential experiments and introduced two distinct types of masking. A *peripheral* one “originating from overlapping on stimuli in the end organ [the ear]” that “can only occur when the two tones excite the same region on the basilar membrane”; a *central* one that is “generally relatively

small and resulting from the conflict of sensations in the brain” and “is probably always present to a certain extent” (Wegel & Lane, 1924). Later on, Pollack (1975) coined the terms *energetic* and *informational* masking as a reference to Wegel & Lane (1924)’s peripheral and central masking. While achieving a precise operational definition of energetic and informational masking across all stimuli and measurement procedures remains challenging (Kidd et al., 2008), energetic masking arises from maskers sharing similar spectrotemporal structure as the signal thus diminishing the availability of signal features and impacting its neural representation as early as at a peripheral level. On the other hand, informational masking generally involves masking effects despite the peripheral auditory system providing sufficient frequency, temporal, and intensity information to the central auditory system to encode both the target and the masker. An intricate interplay between energetic and informational masking arises when the task involves distinguishing relevant speech from irrelevant speech.

1.1.1.2 Speech recognition in speech, a special case of ASA

Among all the crucial functions of the sense of hearing, human listeners heavily rely on the ability to selectively attend to a single talker among multiple simultaneous speakers and effectively follow the communication flow in conversations. This ability plays a pivotal role in numerous everyday social interactions and is, at least for listeners with normal hearing, typically achieved successfully and with relative ease (Kidd & Colburn, 2017). As Miller accurately wrote: “It has been said that the best place to hide a leaf is in the forest, and presumably the best place to hide a voice is among other voices” (Miller, 1947).

The unique challenge presented by speech mask stems from its intricate and dynamic temporal structure, coupled with its striking resemblance to the intended target speech. As highlighted earlier, energetic masking arises when

two competing signals possess similar spectrotemporal and intensity attributes. The task of attending to a speech signal within speech masking becomes demanding due to the shared acoustic characteristics between the target and the interfering mask. One mechanism through which listeners overcome such masking is by taking advantage on the highly fluctuating spectrotemporal pattern of the speech masker. Specifically, in continuous speech, instances of reduced spectrotemporal energy, referred to as *glimpses*, occur repeatedly (Rosen et al., 2013). These glimpses offer opportunities of released masking effects allowing to process the target speech effectively. Notably, in cases where the speech masker consists of a single talker, these instances of masking release are frequent. However, when the number of talkers contributing to the speech mask increases, as is common in real-world scenarios, these masking release opportunities become less frequent. As a result, multi-talker noise, often referred to as babble noise, stands out as the most challenging noise condition for speech perception, closely resembling the complexity of ecological conditions.

Several factors have been shown to influence masking efficiency in speech-in-speech recognition. These factors include the spatial separation between the target and the masking speech (Litovsky, 2005; Yuen & Yuan, 2014), the introduction of a frequency mismatch between the target and masking speech as well as the distribution of formant frequencies (Fitch & Giedd, 1999; Misurelli & Litovsky, 2015; Wightman & Kistler, 2005), mismatches in the language of the target and masking speech (Calandruccio et al., 2016), and temporal onsets mismatches (Hukin & Darwin, 1995). One pivotal factor that significantly impacts the masking effectiveness of speech is the presence of multiple talkers in the background (Rosen et al., 2013; see also Figure 1). In a seminal study in which signal-to-noise ratio (SNR, the ratio of

signal intensity to masker intensity, expressed in dB) was held constant, Miller (1947) observed that accurate identification of isolated words decreased as the number of talkers within the babble masker increased across one, two, four, and six talkers. For nonsense sentence targets, performance decreases with 2 talkers in the masker in contrast to one talker (Freyman et al., 2004). Interestingly, beyond a certain number of talkers in the babble masker, performance seems to rise again, a phenomenon observed between six to eight talkers in Miller (1947)'s research and extending beyond eight talkers up to 512 talkers when using vowel-consonant-vowel targets (Simpson & Cooke, 2005). Similarly, Rosen et al. (2013) documented a marked decrease in sentence recognition between one and two talkers, followed by a modest increase in recognition between two and sixteen talkers.

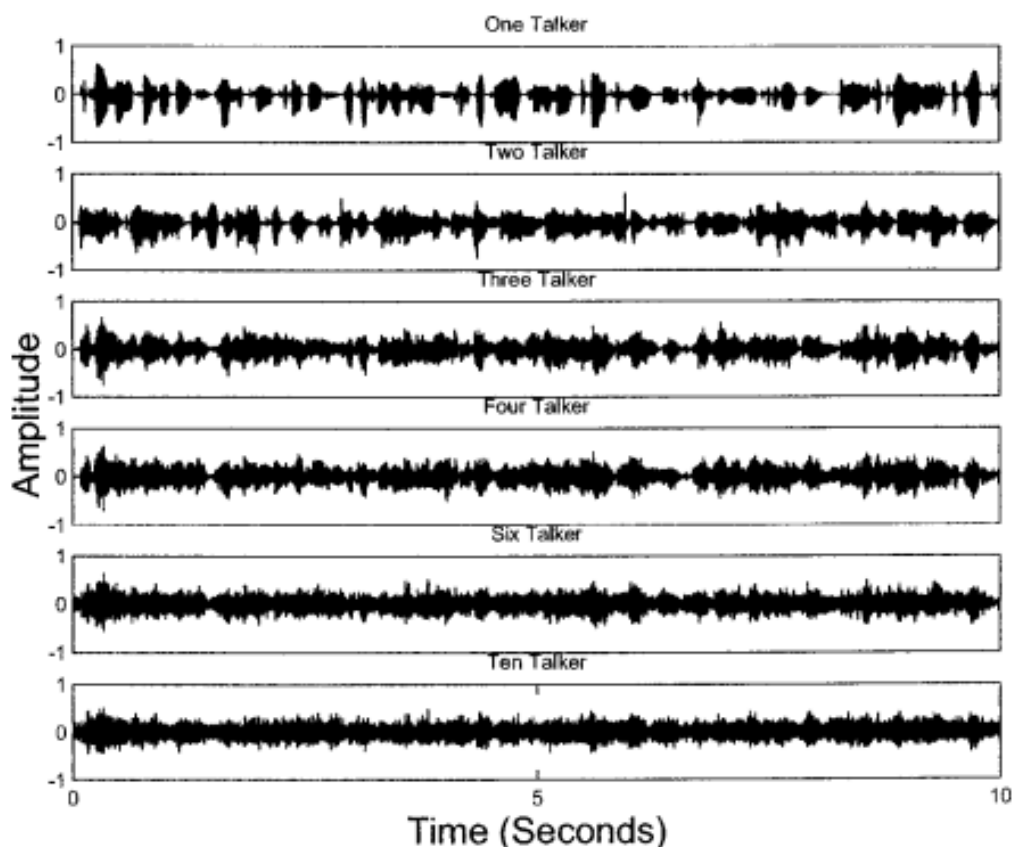


Figure 1: ten-second-long segments of single and multi-talker maskers, from Freyman et al. (2004). Note that the more talkers are present in the masker, the less spectro-temporal fluctuations are available to take advantage of the glimpses to process target speech.

This non-linear shift in performance might appear counterintuitive, given our expectation of a consistent decline in performance as the number of talkers increases. The rationale behind this dynamic rests on the fact that as the count of talkers in the background increases, their overlapping energy starts to fill in the spectral and temporal dips in the masker, diminishing the availability of glimpses to process the target speech (see Figure 1). Thus, we would anticipate that the energetic masking effect of such a background with multiple talkers would exhibit a steady rise with the increasing number of talkers, up until a point where the number of talkers reaches a level where additional talkers have negligible additional impact. In this context, an energetically effective masker would be a broadband noise tailored to match the spectrum of the target speech (e.g. speech-shaped noise) because such a noise has minimal fluctuations that allow for glimpses. This does not seem to be the case as even when dealing with 128 talkers, babble noise remains more challenging than a speech-shaped noise (Simpson & Cooke, 2005). A plausible explanation for the non-linear effect of increased talkers in the speech masker could be attributed to the concurrent increase in energetic masking and decrease in informational masking as the number of talkers increases. The presence of more background talkers could potentially enhance target speech intelligibility by making the background less perceptually similar to the target (Simpson & Cooke, 2005). Likewise, increasing the number of talkers might render individual words in the babble less discernible, thereby reducing lexical interference (Hoen et al., 2007). Consequently, the intricacies of speech recognition within babble noise require a meticulous consideration of the number of talkers in the speech masker, as it involves a complex interplay between energetic and informational masking.

Children face early on the task of learning communication within noisy settings, such as classrooms. Consequently, the research conducted over the last

three decades has been dedicated to understanding the abilities of children to recognize speech in masking conditions. Children are more susceptible to auditory masking situations than adults and their ability to cope with it does not mature until adolescence (Corbin et al., 2016). Most importantly, children's speech recognition is more affected when the masker is also speech than when the masker is steady-state noise (Leibold & Buss, 2019). In the next section, we will review current knowledge about the development of speech recognition in speech maskers throughout school years.

1.1.1.3 Development of speech-in-noise perception

Children display more difficulty than adults in recognizing speech when faced with background sounds, and this trend is consistent across various stimuli and listening conditions (Kidd & Colburn, 2017). Investigations into children's ability to recognize speech in steady-state noise highlight that these differences between children and adults remain noticeable until around 9–10 years of age (for a review on development on speech recognition in steady-state noise, see Leibold & Buss, 2019). Conversely, differences between children and adults seem to be more substantial and persist later (until adolescence) when the masking stimulus itself consists of speech. Hall et al. (2002) used a forced-choice, picture-pointing task to measure the recognition of spondee words (word comprising two equally stressed syllables, e.g., “popcorn”) within the presence of either speech-shaped noise or speech from two talkers. Participants were children aged 5 to 10 years and adults aged 19 to 48 years. On average, children needed an additional 3 dB to match the performance level of adults in the noise masker. In contrast, there was an 8-dB difference between adults and children in the two-talker masker to reach similar performance. Using a similar task to assess recognition of consonant-vowel stimuli differing only by consonants, Leibold & Buss (2013) showed that performance was adult-like at 11 years of

age for the speech-shaped noise. However, in a two-talker masker, performance was lower than adults until the oldest age group (13 years old). Corbin et al (2016) measured children's speech reception thresholds (SRT, thresholds for 50% correct recognition) of monosyllabic words by varying the masker's intensity with an open-set paradigm. Comparably to the previously described studies, children reached adult-like performance by 10 years of age when the masker was a speech-shaped noise, while children's performance increased up until 13 years of age in a two-talker masker. These studies suggest that with a speech masker comprising two talkers, regardless of the stimuli used for recognition and of the task (open/closed set), children's performance reach adult's performance later than for speech-shaped noise.

Only a few studies investigated the development of speech recognition in speech masker with four or more talkers (i.e., babble noise). One of the first study of children's perception in babble noise measured final-word repetition of sentences and showed increased performance from 9-year-olds to 17-year-olds, even between 15 and 17-year-olds, suggesting an even longer developmental trajectory for speech in babble noise than for speech recognition in two-talker maskers (L. L. Elliott, 1979). Using an open-set syllable recognition paradigm in one-, four-, and eight-talker babble noises, Calcutt et al. (2018) showed that children exhibited a significant decrease in speech perception scores from a single-talker babble noise to a four-talker babble noise without significant difference between the four-talker and eight-talker babble noise. Interestingly, dyslexic children displayed lower performance than controls in babble noise and modulated speech-noise, but not in quiet or continuous speech-noise. Overall, speech perception in speech seem to be a long maturation process along childhood and adolescence. More specifically, despite the scarcity of studies on the developmental trajectory of speech perception in babble noise, it seems to

be the most detrimental situation that children encounter during development, reaching its full development only in late adolescence.

When designing a speech-in-noise test suitable for children, the choice of the masker is only one aspect to consider. While many prior studies have employed open-set response paradigms, these can be limited in their applicability to children. For instance, the need for clear verbal articulation poses challenges for hearing-impaired children, or with speech sound production disorders (Cabbage & Hitchcock, 2022). To address these challenges, alternative forced choice tests (nAFC) that involve picture-pointing responses, where the child selects an answer from multiple options, have proven valuable in pediatric assessments (Vickers et al., 2018). Moreover, word recognition is influenced by psycholinguistic factors like word familiarity and occurrence frequency (Howes, 1957; Savin, 1963). Words with high occurrence frequency and fewer phonological neighbors are generally easier to recognize and affect open-set word recognition in both adults and children (Krull et al., 2010; Meyer & Pisoni, 1999). Closed-set tests, in contrast, present potential answers, reducing the reliance on their mental lexicon and reflecting auditory abilities more directly. This testing format can help mitigate unwanted variability in signal processing, especially in children whose mental lexicon organization varies with age. Such independence from vocabulary knowledge is a recommended criterion for speech perception tests in children (Mendel, 2008). These considerations are explored in the introduction of the first study of this PhD (section 2.2).

Children appear to rely on many of the same acoustic differences between target and masker than adults. Children as young as 4-years old are able to take advantage of the spatial separation between target and masker (i.e., spatial masking release, SMR) for speech recognition in speech-shaped noise, one-, and two-talker noise (Litovsky, 2005). However, Brown et al. (2010) showed that

the ability to benefit from SMR in sentence recognition in a two-talker noise was poorer than adults up until 14 years of age. Another acoustic cue that children are able to take advantage of is the frequency difference between target and masker. When asked to recognize dissyllabic words in a two-talker masker, children showed poorer performance when target and masker were sex-matched than when they were mismatched (Leibold et al., 2018). However, many acoustic cues can differ between talkers of different sexes like fundamental frequency, formant frequencies and phonation types (Leibold & Buss, 2019). Varying only mean fundamental frequency between target and masker revealed that benefits from fundamental frequency separations increased with age (7- to 15-year-olds) and adult-like benefit was reached at 13 years of age (Flaherty et al., 2019). Moreover, by investigating individual data, Flaherty et al (2019) showed that children below 7 years of age did not benefit from the fundamental frequency separation, suggesting that younger children might need additional acoustical cues to isolate target and masker. As mentioned in section 1.1.1.2, adults rely on glimpses during the minimal spectro-temporal energy of the masker to isolate target speech. Buss et al (2017) measured SRTs in children and adults in one- or two-talker masker with or without an additional speech-shaped noise. The reason behind evaluating performance under the influence of speech-shaped noise was to investigate how the presence of noise could obscure low-level speech cues that are typically present during the glimpses of the speech masker's envelope. The inclusion of speech-shaped noise to speech masker led to increased SRTs in children and adults and its impact was more pronounced with single-talker masker compared to a two-talker masker (see Figure 2). Furthermore, the influence of speech-shaped noise on SRTs showed a correlation with the age of the listener: noise had a relatively milder effect on young children compared to older children and adults. These results suggest

that children are not as proficient as adults in identifying speech based on brief glimpses. Overall, there is a maturation in children on the use of spatial cues, acoustic cues, and momentary glimpses – all of which are elements that adults depend on to separate speech from speech. These factors seem to collectively play a role in the progressive increase of speech perception in speech as children develop. Beyond those factors, immature allocation of attention has been suggested to play a key role in children’s pronounced speech-in-noise difficulties (Leibold & Buss, 2019). The next section will be dedicated to the presentation of selective attention and its development, including its role in ASA and speech-in-noise perception.

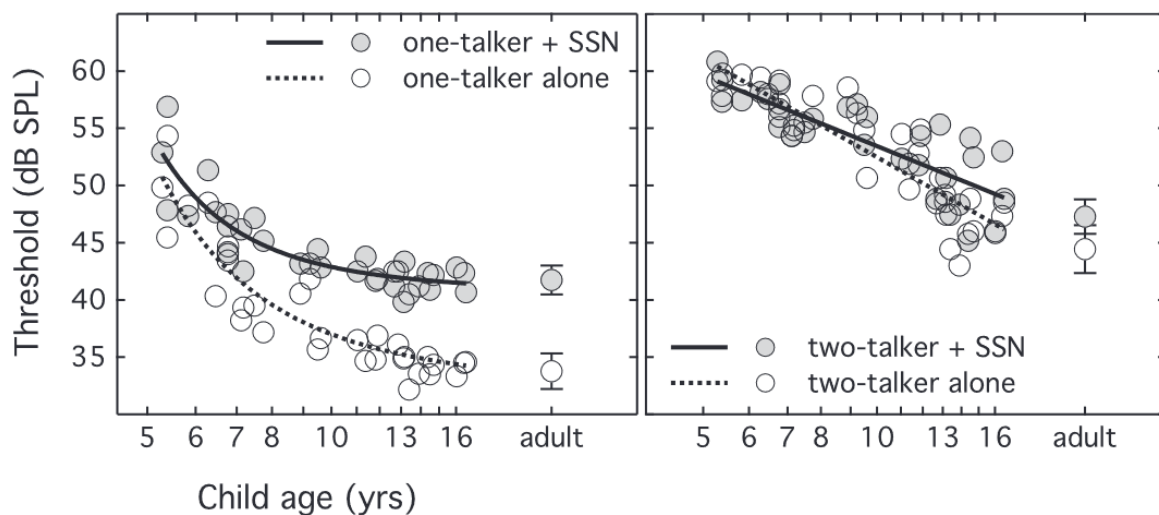


Figure 2: developmental trajectory of speech recognition (measured as speech reception thresholds, SRTs) in one- (left panel) or two-talker (right panel) maskers, with (filled circles) or without (plain circles) additional speech-shaped noise (SSN) in children from 5 to 16 years old and young adults. Lines indicate data fits. Children reached adult-like performance at 10-12.9 years of age in the one-talker masker, while adult-like performance was not reached until 16.1-16.8 years in the two-talker masker. Additionally, SSN has a greater detrimental effect in one-talker masker than in two-talker masker, suggesting that SSN interfered with glimpses in the one-talker masker, and that glimpsing plays a smaller role when more than one talker is present in the masker. Figure from Buss et al. (2017).

1.1.2 Auditory attention

Attention plays a vital role in most, if not all, information processing. It involves the selection of internal and external stimuli that deserve further processing and response. This selection process is influenced not only by the characteristics of the stimuli but also by the individual's interests, motives, and cognitive strategies (Fritz et al., 2007). In many communication scenarios, multiple acoustic sources of information are available for auditory processing. The ability to focus one's attention on one auditory object while disregarding others in the background is referred to as auditory selective attention. Auditory selective attention mechanism is believed to play a role in various stages of ASA: from *bottom-up* segregation of input's spectrotemporal features to *top-down* refining of sound stream segregation (Sussman, 2017). We will thus focus in this section on a brief presentation of auditory selective attention followed by an overview of its development throughout childhood.

1.1.2.1 Auditory selective attention

One of the most famous complex sound environments is the “cocktail party” situation in which we are exposed to multiple streams of sounds (as in a classroom with the teacher's voice, whispered conversations, chairs moving, school accessories falling with a sudden clatter, etc., Cherry, 1953). The investigation of this scenario became prominent in the 1950s, particularly with Broadbent's work (1958), which compared the brain to a computer with limited resources, unable to process all streams of sounds simultaneously (see Figure 3). Broadbent proposed the “early-selection” model, suggesting that attention filters inputs at a lower level where acoustic properties are processed, allowing only relevant information to pass through for further processing. This idea originated from dichotic listening experiments, where participants focus on one auditory stream in one ear and ignore the other in the other ear. While initial

results seemed to support the complete blocking of stimuli in the ignored ear, later studies showed more nuanced findings, with participants being able to tell when their own name was presented in the ignored ear, indicating that some processing of the unattended information still occurs at a certain level (Moray, 1960). Treisman (1960) proposed a modified version of Broadbent's model, suggesting that irrelevant inputs are attenuated rather than entirely blocked during attentional filtering. Subsequently, "late-selection" models for selective attention emerged, proposing that filtering occurs at later processing stages, where irrelevant stimuli are extracted during identification or semantics stages based on a comparison with targeted information (J. A. Deutsch & Deutsch, 1963; Duncan, 1980). Kahneman (1973) introduced an alternative limited-capacity model, proposing that attention is a finite resource, and the complexity of the input affects the level of processing of the unattended information.

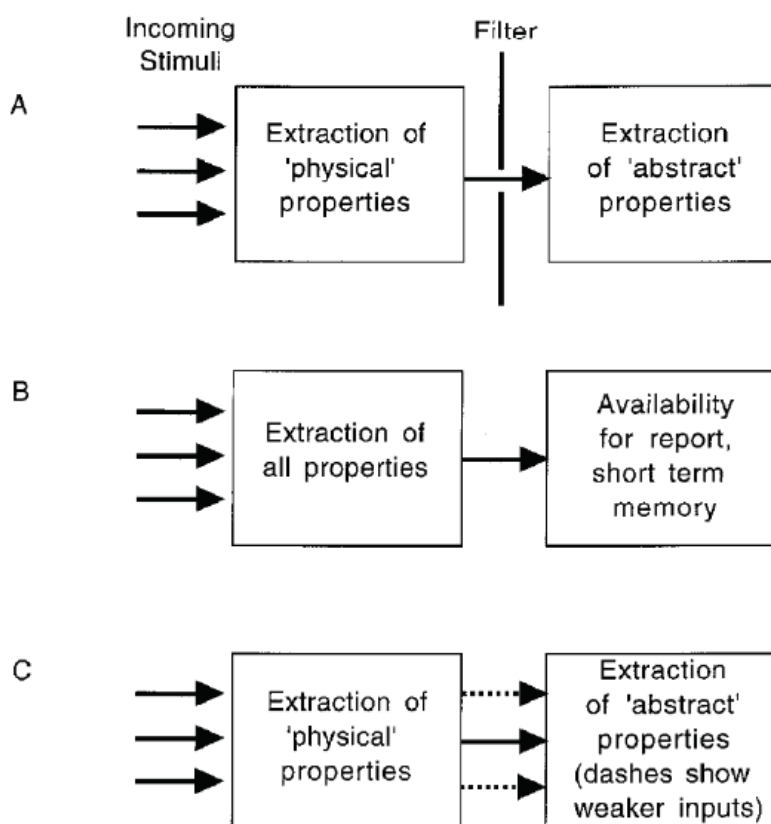


Figure 3: Representations of models of selective attention. (A) Broadbent's "early-selection" model (1958); (B) the "late-selection" model (J. A. Deutsch & Deutsch,

1963; Duncan, 1980); (C) Treisman (1960)'s attenuation version of Broadbent's model. Figure from Driver (2001).

In the context of ASA (i.e., segregating multiple sound sources in perceptual streams, see section 1.1.1.1), whatever the combination of acoustical cues available to segregate them, we accomplish this complex task of selectively attending to a single stream in everyday life and in various acoustical environments (Fritz et al., 2007). Given the complex nature of auditory environments, the significance of a sound event can stem from the scene itself (*bottom-up*), like a prominent sound such as a gunshot that naturally grabs attention. Or it can be shaped by task-oriented objectives and expectations (*top-down*), such as when engaging in a conversation among competing auditory sources (Kaya & Elhilali, 2017). While the significance of bottom-up salience is undeniable, voluntary top-down attention holds the primary responsibility for selecting foreground elements over background and for shifting the focus of attention towards various features, objects, or streams of importance within the auditory environment (Fritz et al., 2007). Because attention influences nearly every processing stage of multiple sound source, isolating its precise impact at each step is challenging. Various methodologies have been employed to assess its influence (Werner et al., 2012). However, delving into a comprehensive review of five decades of attention-related literature exceeds the scope of this PhD. Instead, the next section will provide a brief overview of the current understanding of attention's developmental aspects in children.

1.1.2.2 Development of auditory selective attention

Research on the development of auditory selective attention remains relatively underexplored due to the absence of appropriate behavioral paradigms to effectively separate attentional processes from other factors

(Leibold & Buss, 2019). In speech processing contexts, selective attention is typically measured by quantifying the errors and intrusions from distracting speech, presented either in the ipsilateral or contralateral ear of the target speech, which listeners must ignore. Notably, these paradigms bear a striking resemblance to those employed for evaluating speech recognition within speech. This similarity highlights the complex interplay between selective attention processes and speech-in-noise recognition, making their differentiation challenging.

Using a shadowing paradigm (Cherry, 1953), in which listeners are instructed to immediately repeat the target message while ignoring a competing distractor message, Doyle (1973) observed that children encounter greater challenges in shadowing the intended message in contrast to adults, often exhibiting responses that include multiple interferences from the competing speech message. Moreover, older children exhibit superior performance and fewer errors compared to younger children. This pattern of development implies an ongoing refinement of selective auditory attention throughout childhood. More specifically, Doyle (1973) observed that younger children displayed a higher tendency to report words or partial words from the distractor speech in comparison to their older counterparts. In light of these findings, the author suggested that younger children exhibit a poorer ability to filter out irrelevant auditory information than older children (i.e., poorer inhibition processes, higher distractibility, poorer sustained attention). More recently, Wightman and Kistler (2005) explored auditory selective attention in listeners from 4 to 16 years old and adults. Using a closed-set speech recognition task, they presented participants with a target speech message along with a distracting speech message in the same ear. Two additional conditions were introduced with either another distracting speech message or speech-shaped noise to the

opposite ear. Children's performance was consistently inferior to that of adults across all conditions, showing improvements in development up until 13 years of age. The introduction of a competing distractor sentence presented to the opposite ear affected performance for listeners across all age groups. Analysis of error patterns revealed age-effects in the ability to disregard speech presented to the contralateral ear with higher intrusions from the contralateral distractor speech for younger children. Leibold & Buss. (2013) also reported more segregation difficulties in 4- and 7-years old children as compared to adults in a two-talker masker. Overall, these studies suggest that children display more difficulty in segregating sound streams resulting in more susceptibility to intrusions from maskers. These results are in line with literature on the attentional distractibility of children that show that the observed high distractibility of young children is the result of reduced sustained attention capacities (Hoyer et al., 2021) that are particularly at play in the context of speech recognition in adverse conditions (Gomes et al., 2000).

Attention thus plays a crucial role in selecting auditory relevant information. In everyday communication situations, processed and identified auditory object can be encoded and stored in memory to be used in relevant short- or long-term goal-oriented behaviors. Attention continues to play a crucial role in these working/short-term memory (WM/STM) processes that refer to the ability to maintain (STM) and manipulate (WM) information during a short period of time (see section 1.1.3). An attentional component has been consistently described as necessary to maintain information in short-term storage (Baddeley, 2010; Cowan, 1998) and to perform concurrent tasks while maintaining information (Barrouillet & Camos, 2007). Indeed, the influence of attentional processes have been described in all stages of STM processing (i.e. encoding, retention, and retrieval, see section 1.1.3.3) in the visual and auditory

modality (e.g., Banbury et al., 2001; B. L. Elliott & Brewer, 2019; Lozito & Mulligan, 2006; Park et al., 1989). More recently, shared cognitive resources have been evidenced between auditory selective attention and STM by simultaneously measuring the ability to ignore distractors (auditory selective attention) and to compare two non-verbal auditory sequences (STM, Blain et al., 2022). The next section will be dedicated in presenting current knowledge about auditory STM, its relation with working memory models, and its developmental course.

1.1.3 Auditory short-term memory

In the following sections, we will first present an overview of the theoretical account of memory, followed by a presentation of various working memory (WM) models, and an examination of the short-term memory (STM) construct in relation to these models. Subsequently, we will delve into the specifics of auditory STM for both verbal and musical material, and finally, we will explore its cognitive development across school years. Note that this section is more elaborated than the two previous ones, as auditory STM is the central cognitive process under scrutiny in three of the four experimental chapters of this PhD.

1.1.3.1 Memory

Within the domain of cognitive psychology and neuroscience, memory stands out among various complex processes such as emotion, attention, learning, decision making etc., receiving extensive interest. This cognitive ability has been recognized as crucially influencing our daily life activities, shaping our identity, knowledge, and perceptual abilities. While the ubiquitous role of memory in human cognition is widely acknowledged, its organization has historically been a subject of debate, revolving around two main theories.

One perspective considers memory as a homogeneous entity, while the other views it as an organization of multiple interacting systems.

The first reported experimental investigation of memory was the observation by Hermann Ebbinghaus that when learning lists of non-sense syllables, recollection of these syllables was better according to the number of repetitions during the learning process (Ebbinghaus, 2013/1885). Later on, on the basis of introspective analysis, William James introduced the distinction between a limited storage of information he called *primary memory* and an unlimited storage of information referred as *secondary memory* (James, 1890). Subsequently, Drachman and Arbib (1966) showed that patients with bilateral hippocampal damage, although they retained their digit span abilities, were unable to learn new words. This significant finding prompted them to introduce the terms *short-term* and *long-term* memories as designations for the primary and secondary memory constructs introduced by William James (1890). This dualistic view of memory is supported by neuropsychological studies in brain-damaged patients that have shown that specific behavioral impairments in either short or long-term memory processing occur depending on the location of the lesion. For instance, Milner (1966)'s report indicated that patients with damages in temporal lobes and hippocampus displayed difficulties in learning and remembering new material (verbal or visual), while maintaining normal STM abilities, as measured by the digit span task. In contrast, another study by Shallice and Warrington (1970) revealed that patients with damage in the left hemisphere's perisylvian region exhibited near-normal long-term memory abilities but significant impairments in STM capacities (digit spans limited to one or two items). Based on these observations, the prevailing notion in the literature was that memory could be explained by the existence of both a short-term store and a long-term store. Atkinson & Shiffrin (1968) further expanded

on this concept by proposing a memory model based on the duration for which information remains in the system (Figure 4). According to their model, external information enters the system through evanescent sensory memories and then progresses into a limited capacity short-term store. Once information is held in the short-term store for a sufficient duration, it can be transferred to the long-term store for more permanent retention. Since these seminal studies, research in cognitive psychology has rather consistently considered memory as an arrangement of multiple systems, despite the attempts of other memory models to define it as a unitary system (Ranganath & Blumenfeld, 2005). In the present PhD work, we will also adopt this perspective, as it aligns with the prevailing literature on auditory STM, which has predominantly been based on the notion of a multi-component memory system.

Since the groundbreaking work of Atkinson & Shiffrin (1968), researchers have identified multiple distinct forms of memory, including three main categories. (1) Sensory memory, also known as echoic memory in the context of auditory processing, allows the retention of information for an extremely brief period, usually less than 10 seconds in the auditory modality, and is closely associated with the processing of information within the perceptual system (Cowan, 1984). (2) Short-term and Working memories align with what James (1890) referred to as “primary memory”. STM is capacity-limited and lasts brief period of time but can be maintained almost indefinitely through maintenance strategies (e.g., rehearsal, attentional refreshing, etc.). On the other hand, WM involves not only the maintenance of information but also the manipulation of information stored in memory. The precise distinction between WM and STM is a topic of ongoing debate (Aben et al., 2012; Baddeley, 2012; Cowan, 2008; D’Esposito, 2007), but there is a general consensus that they represent systems dedicated to maintain (and manipulate, in the case of WM) information over a

Auditory short-term memory

short period of time (several seconds, Baddeley, 2010; Cowan, 2008). (3) Long-term memory is a highly expansive memory system capable of storing information for extended periods (from hours to a lifetime, Camina & Güell, 2017). While these memory systems have been defined since the work of Atkinson & Shiffrin, the precise boundaries and characteristics of each system are still under discussion among researchers, with ongoing debates in the field (Baddeley, 2012; Camina & Güell, 2017; Cowan, 2008). Nonetheless, these distinctions provide a valuable framework for understanding the diverse aspects of memory and their roles in cognitive activities.

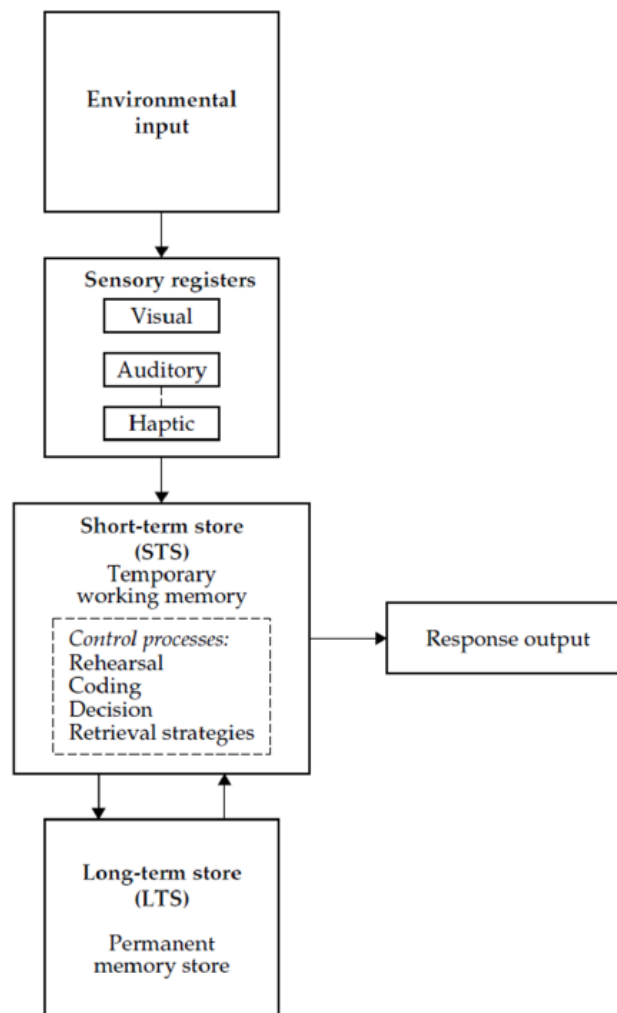


Figure 4: The memory model proposed by Atkinson & Shiffrin (1968) outlines a sequence of memory storage systems through which information progresses. It starts with input from the environment, goes through temporary sensory buffers, and

subsequently enters a restricted-capacity short-term memory store. This store, according to Atkinson and Shiffrin (1968), serves as a functional working memory responsible for managing the transfer of information to and from long-term memory. This dynamic process holds significant importance in facilitating learning and cognitive functions overall. Figure from Atkinson & Shiffrin (1968).

1.1.3.2 Working memory

Over the past decades, the distinction between STM and WM has remained ambiguous due to various inconsistencies in their usage (Baddeley, 2012; Cowan, 2008). Some researchers have used “short-term memory” to describe the simple temporary storage of information (maintenance), while “working memory” has been used to encompass both the maintenance and manipulation of information (Engle et al., 1999). Early studies proposed that WM is integrated within short-term stores (Atkinson & Shiffrin, 1968). However, more recent models have regarded STM as a sub-component of WM (Cowan, 2008). This lack of clear delineation has contributed to ongoing debates and challenges in precisely defining the boundaries between these two constructs. We will present in this section the main models of WM.

WM “refers to a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning” (Baddeley, 1992). This term has become highly influential in cognitive psychology since later theoretical accounts of the dynamic processing of information in memory have raised at least two problems in Atkinson & Shiffrin (1968)’s memory model. (1) Their model posits that the longer items remain in the short-term store, the better they will be encoded in the long-term store. However, this idea has been contradicted by studies that demonstrate no relationship between the duration of an item's stay in short-term storage or the number of overt rehearsals and its subsequent recall performance (Craik & Watkins, 1973). (2) Their model

assumes that short-term storage is an obligatory step for long-term storage. This notion was challenged by studies that observed a double dissociation between short-term and long-term storage in patients who were impaired in one storage type while remaining unimpaired in the other (Milner, 1966; Shallice & Warrington, 1970). Given these criticisms, new models have proposed alternative theoretical accounts for working/short-term memory (Baddeley & Hitch, 1974; Cowan, 1988, 1998).

Baddeley and Hitch (1974) initially introduced a three-component system consisting of a central executive, a phonological loop and a visuospatial sketchpad. The phonological loop “is probably the best developed component of the working memory model” (Baddeley, 2000). It comprises a short-term storage subject to rapid temporal decay for which linguistic inputs have automatic access. This component can be further divided into two parts. (1) A passive storage component (phonological storage) responsible for maintaining auditory or speech-based information for a brief period, typically a couple of seconds and (2) an active rehearsal mechanism aimed at preserving information for longer durations through articulatory rehearsal processes, such as subvocal speech. The visuospatial sketchpad is assumed to allow the active storage and manipulation of visual and spatial information. Its functioning is beyond the scope of the current PhD work as it does not concern the auditory modality. The central executive is believed to serve as an attentional controller for the visuospatial and phonological subsidiary systems of working memory. Its role integrates several processes such as dividing and switching attention, along with facilitating the connection between working memory and long-term memory (Baddeley, 1996). Subsequently, as a result of observations of various phenomena that the original model could not fully explain, they proposed the inclusion of a fourth component: the episodic buffer (Baddeley, 2000, see Figure

5 for a representation of the full model). This component acts as a temporary storage system that integrates information from different sources and modalities, functioning as a form of “conceptual short-term memory” (Camina & Güell, 2017).

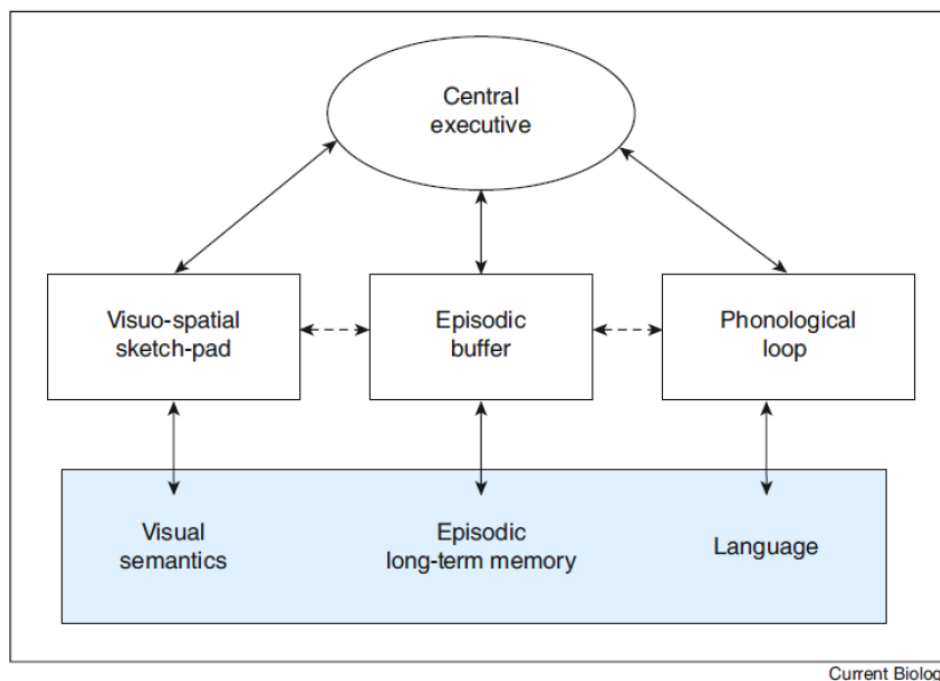


Figure 5: *Baddeley & Hitch's working memory model, from Baddeley (2010).*

Over the past decades, theoretical accounts of WM have shifted from multi-store models and structural approaches to more dynamic and process-oriented models of WM (Barrouillet & Camos, 2007; Cowan, 1988; Engle et al., 1999). These theories have shifted the conception of WM from being viewed as a separate structure or a set of cognitive structures to considering it as the activated portion of long-term memory. By adopting this perspective, they have focused on the functional aspects of cognition, exploring how activation is generated, sustained, or inhibited, and how cognitive processes use the activated information. Consequently, these models are more concerned with understanding “how it works” rather than “how it is organized”. One such model representing these dynamic approaches is the one proposed by Cowan (1988) that emphasizes the role of attention in managing activated information

(Figure 6). It proposes that WM capacity is limited, and the activated information is subject to rapid decay if not continuously in the focus of attention. If in theory, this model seems far removed from Baddeley's, the "limited capacity attentional focus" is similar to the central executive component and some of the evidence upon which the two models are built are the same but led to different taxonomies.

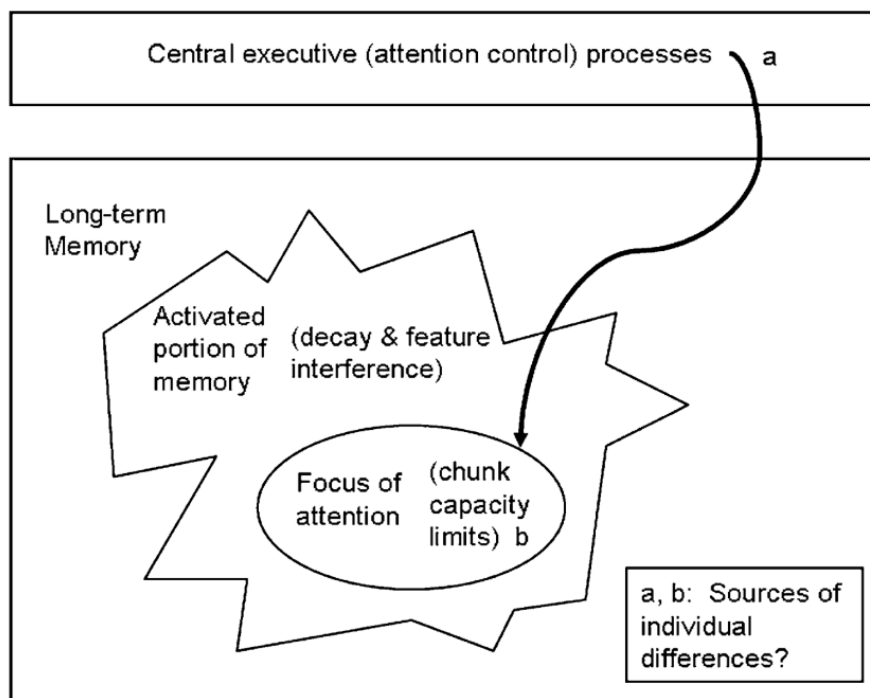


Figure 6: a depiction of Cowan (2008)'s theoretical framework of working memory. Cowan's theory challenges the notion that the 'contents of working memory' are stored in distinct dedicated buffers. Instead, it suggests that working memory consists of information currently within the focus of attention. Cowan's model organizes working memory into a hierarchy that includes long-term memory, the currently activated subset of working long-term memory, and the specific subset of activated memory that is in the focus of attention. Figure from Cowan (2008).

WM has been extensively explored in cognitive psychology and neuroscience. Through studies involving both patients and typical individuals, researchers have identified several sub-components and dynamics of the system responsible for maintaining and manipulating information in the short-term period. However, the line of demarcation between STM and WM appears to

lack clarity in contemporary research. Notably, some studies have employed the term "working memory" for tasks that involve only maintenance processes (Schulze & Koelsch, 2012) or some researchers would like to reserve the term "working memory" only for the attention-related aspects of STM (Engle, 2002). To provide a more precise definition, the following section will introduce the concept of STM with regard to WM models and present the main paradigms used for its investigation.

1.1.3.3 Short-term memory (STM)

Historically, the term "short-term memory" originally referred to James' concept of "primary memory" (1890). However, this definition was soon deemed too vague as "primary memory" could also be associated with sensory memory (Cowan, 2008). Later, Atkinson and Shiffrin (1968) defined STM as a temporary store capable of holding a limited amount of information for a short duration. As discussed in the previous section, Baddeley and Hitch (1974) proposed a multi-component system that rejected the notion of a singular short-term store. Instead, they introduced the term "working memory" to describe the entire system, along with its subcomponents. One approach to distinguish between short-term and working memory is to consider that the subsidiary systems proposed by Baddeley and Hitch (phonological loop, visuospatial sketchpad) might function as short-term storage components, with these components being manipulated by central executive processes. Cowan (1988) adopted a multi-component perspective similar to Baddeley and Hitch's approach. However, he did not commit to specific components as they did. Instead, he characterized the fundamental divisions of working memory as short-term storage components, which encompass both activated memory and the focus of attention within it, along with central executive processes responsible for manipulating stored information (Cowan, 2008). In Cowan's

view, Baddeley and Hitch's phonological loop and visuospatial sketchpad would be regarded as just two among various aspects of activated memory (Figure 6). These aspects are susceptible to interference to varying degrees, depending on the similarity between features of the activated information and interfering information sources. In summary, the two main types of WM models described earlier both concur that STM is a component embedded within working memory.

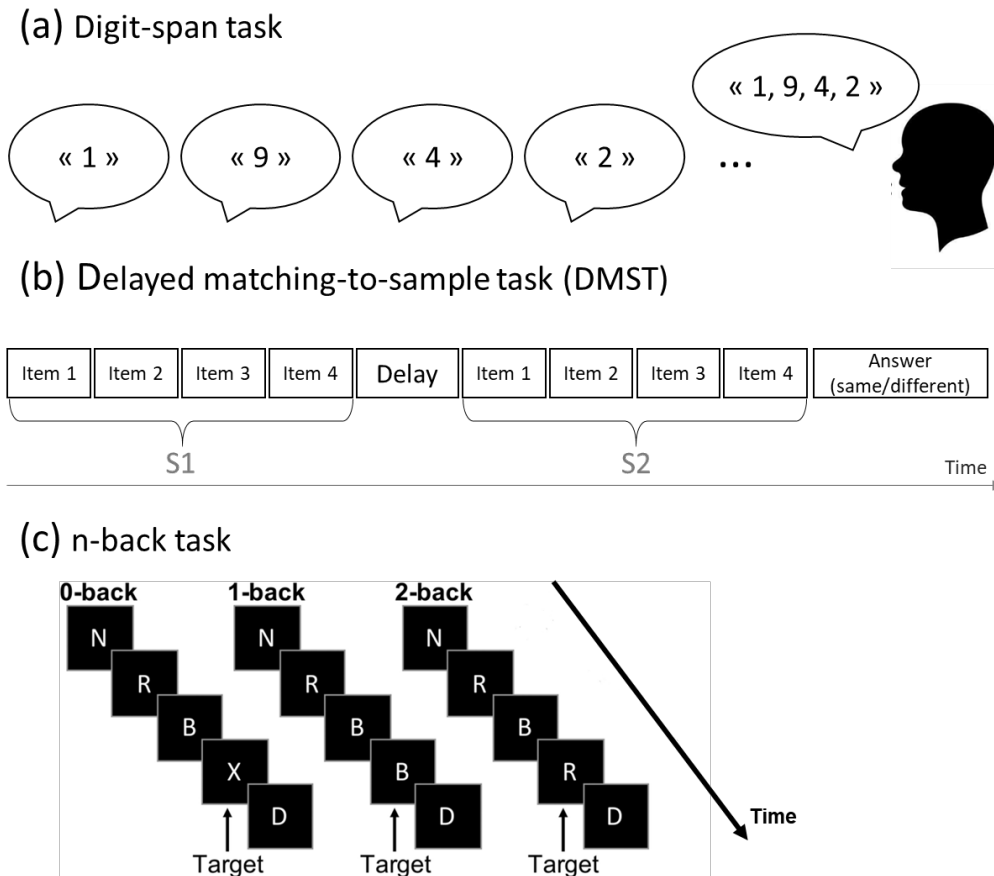


Figure 7: (a) Example of a digit-span task (serial recall task) where the participant is asked to repeat a sequence of items in the order it was presented. (b) Schematic representation of a typical delayed match-to-sample task (DMST), here with sequences of four items. When the S2 sequence is different from S1, either a new item can be introduced, or two items can be switched. (c) Schematic representation of a typical n-back task. Stimuli are presented sequentially auditorily or visually. In the 0-back task, the participant is instructed to produce a response whenever a target (in this example the letter X) appears. In the 1-back task, the participant has to produce a response when a target (here the letter B) was present in the previous presentation. In the 2-back task, the participant has to produce a response when the target (here the letter R) was present two presentations before. Figure from Costers et al. (2020).

Note that the (a) and (b) paradigms specifically measure STM processes as only maintenance of information is needed, whereas (c) measures WM processes as it entails the maintenance and the updating of upcoming information.

In cognitive psychology, the investigation of STM has often focused on three essential aspects of any memory system: Encoding, the processes involved in registering the information; Storage (or Retention or Maintenance), responsible for maintaining information over time; and Retrieval, the process of accessing information through recall or recognition. Once information is processed by the perceptual systems, it needs to be encoded in a form that the cognitive system can handle. This encoding phase is typically studied by varying how the material is processed (e.g., modality of presentation, timing of stimulus presentation, inter-stimulus interval etc.) and modifying the nature of the material to be encoded. For instance, encoding a characteristic of a word results in poorer subsequent recall or recognition compared to encoding it in terms of meaning (Baddeley, 1992). After the encoding step, the information is stored (storage, retention, or maintenance). Retention is measured through the rate of loss of information (forgetting). Retention capacities vary depending on the nature of memory stores, including duration (how long the memory lasts), capacity (how much can be stored at once), and the type of information held. Finally, retrieval involves extracting information from storage. STM retrieval can be investigated using two main methods: 1) recall, where the participant reproduces stimulus items, and 2) recognition, where the participant determines whether items were presented during the encoding phase (yes/no recognition) or selects the previously presented item from a set of alternatives (forced-choice recognition).

For recall methods, the participants can be asked to report items in the order presented (serial recall, see Figure 7a) or in any order (free recall). For both types of procedures, the probability of correctly recalling an item is

influenced by its serial position in the presentation. Typically, recall is accurate at the beginning, decreases throughout the list, and then improves toward the end. The recency effect is thought to reflect the use of a short-term store where recently presented items are still available at the time of recall. The primacy effect is believed to arise from an interplay of attentional, rehearsal and memory capacity factors. Note that recall paradigms require participants to verbally or manually reproduce the items presented during the task. Recognition methods don't require the participant to reproduce the items presented during the tasks but depend on the participant's response strategy and degree of caution. For example, in a yes/no categorization task where participants decide if an item was presented during the encoding phase (see Figure 7b), if a participant responds "yes" to every stimulus, their categorization scores may not reflect memory processing accurately. To address this, various analyses consider different degrees of caution among participants, including signal detection theory. With such a method two measures are taken into account: the sensitivity index (d'), representing the hypothetical strength of the memory trace, and the response criterion (c), representing the participant's level of caution or response bias (Macmillan & Creelman, 2005). Recall and recognition tasks specifically target STM processes (see Figure 7a/b) as they only require maintenance of information in contrast with n-back tasks (see Figure 7c) that target WM processes that require maintenance and manipulation of information. In the upcoming section, we will present the current state of knowledge regarding auditory STM for musical and verbal material and we will advocate the use of recognition methods to investigate these two types of materials in a comparable manner.

1.1.3.4 Auditory STM for verbal and musical material

The majority of research on auditory STM has predominantly focused on verbal material, such as words, syllables, and phonemes. However, more recently, there has been growing interest in investigating memory mechanisms related to other types of auditory material, such as music (tones, timbres, and rhythms, McKeown & Wellsted, 2009; Mercer & McKeown, 2010; Schulze et al., 2012; Schulze & Koelsch, 2012; Schulze & Tillmann, 2013; Talamini et al., 2022). Auditory STM for tones is a crucial element in music perception and production, but it also plays a significant role in speech processing and comprehension. In the subsequent section, we will provide more detailed insights into the current understanding of the mechanisms supporting auditory STM for both verbal and musical material. From this point forward, when discussing musical material, we will refer to tones (of varying pitch) unless otherwise specified.

1.1.3.4.1 Verbal STM

It is now widely accepted that verbal STM can be effectively supported by the phonological loop proposed by Baddeley and Hitch (1974). The phonological loop has been defined based on several behavioral effects (the phonological similarity effect, the irrelevant speech effect, the word-length effect, and the articulatory suppression effect) that indicate that verbal material is stored in STM through its transformation into a phonological code. The *phonological similarity effect* refers to the detrimental effect of sound similarity between to-be-remembered verbal items whereas visual or semantic similarity show very little effects (Baddeley, 1966). The *irrelevant speech effect* refers to the fact that when participants are exposed to irrelevant speech during serial recall tasks, their retention performance tends to be poorer compared to when they listen to pulsed noise (Salamé & Baddeley, 1987). This finding suggests

that the irrelevant speech effect may be attributed to the obligatory access of phonological information (but not other auditory information) during the retention process. The *word-length effect* refers to the phenomenon where shorter words are found to be more easily remembered and recognized compared to longer words (Baddeley et al., 1975). This has confirmed the existence of rapid decay of auditory memory traces in the phonological loop and the existence of an articulatory rehearsal loop to prevent such decay (e.g., with longer words, as the articulatory cycles increase, the likelihood of a particular item within the store undergoing decay grows higher). The *articulatory suppression effect* refers to the fact that when typical individuals are tasked with continuously repeating an irrelevant sound (or naming months of the year) during maintenance of verbal items, their performance is negatively affected. This indicates that articulatory rehearsal processes are essential mechanisms for phonological storage.

Moreover, it is also well acknowledged that verbal STM does not operate in isolation but rather functions within the context of a complex cognitive system. Numerous studies have investigated the potential impact of various processes on verbal STM, especially in children in relation with language development (see section 1.1.3.5). One such process is knowledge stored in long-term memory and several studies have demonstrated its contribution to verbal STM. For instance, familiar words are better recalled than nonsense syllables (Hulme et al., 1991). Long-term knowledge's influence is also evident in other findings, such as the phonotactic frequency effect, where sound sequences with higher probabilities of occurring in the lexicon are better recalled (Pickering et al., 2001); the word frequency effect, where more frequently used words are better recalled (Hulme et al., 1997); and the imageability effect, where words that are easier to form a mental image of are better recalled (Ballot et al.,

2021). These findings collectively highlight the significant impact of long-term knowledge on verbal STM processes. While these mechanisms are becoming better understood for verbal information, it remains an open question whether other inputs, such as music, are supported by the same processes. One possibility is that non-verbal information is recoded as verbal information to be processed by the phonological loop. On the other hand, some researchers suggest the involvement of additional specialized mechanisms, leading to the concept of a “tonal loop” (Berz, 1995; Pechmann & Mohr, 1992). We will explore this issue in the upcoming section.

1.1.3.4.2 Musical STM

Seminal studies have suggested that a specialized subsystem might be responsible for the temporary storage of tonal pitch. Deutsch (1970, 1974) conducted experiments in which participants had to judge if two tones with identical or different pitches with a 5-second retention delay in-between were the same or different while interfering speech or tones were played during the retention delay. The results showed that interfering tones significantly disrupted pitch memory performance, while spoken numbers had little effect, leading to the suggestion of a specific system supporting memory for pitch. Salamé and Baddeley (1989) conducted a series of experiments using serial recall of verbal material and interfering vocal or instrumental music. While they observed detrimental effect of both vocal and instrumental music on untrained participants, they observed detrimental effect of vocal (but not instrumental) music on trained participants. The authors suggested that while speech and music are processed in a common working memory system, both might recruit different sub-systems (or loops).

As a result of these studies, a model of WM has been proposed to include an additional subsidiary system (or short-term store) known as the tonal loop. This component was believed to facilitate the processing and storage of musical stimuli separately from the verbal or visuospatial components (Berz, 1995; Pechmann & Mohr, 1992). In line with this proposition, equivalent behavioral effects that helped define the phonological loop (see previous section) have been observed for musical material:

- Williamson, Baddeley et al. (2010) have demonstrated an equivalent of the *phonological proximity effect* that they termed *pitch proximity effect*. The authors tested participants' STM with either a recall paradigm (where participants reproduced a sequence of tones with a visual grid) or a recognition paradigm (where participants had to indicate if two tone sequences were identical or different). For both tasks, the tone sequences were either proximal sequences (tones with close pitches) and distal sequences (tones with higher pitch differences). For both paradigms, non-musician participants displayed an effect of pitch proximity (i.e. decreased performance with increased pitch proximity).
- Evidence for an equivalent of the *irrelevant speech effect* has also been observed for musical material. Three studies showed that introducing pitch interferents during the silent retention delay of a musical STM recognition task (e.g., comparison of two tones or two sequences of tones, separated by a retention delay) leads to decreased performance when comparing to no pitch interference (Gosselin et al., 2009; Williamson, McDonald, et al., 2010) or speech interferents (D. Deutsch, 1970). These studies show a specific interference effect for musical STM, parallel to the irrelevant speech effect observed for

verbal STM. Furthermore, Semal et al. (1996) showed that performance in a two-tone comparison was better when the difference in pitch between interfering sounds and to-be-remembered tones was higher than when it was small, for speech as well as for nonspeech interfering sounds. This suggests that regardless of the nature of the interfering sound (speech or non-speech), pitch information interferes with musical STM.

- Several studies have reported an effect of the length of musical sequences during STM retention, similar to the *word-length effect* observed for verbal material (e.g., Croonen, 1994; Schulze & Tillmann, 2013), suggesting the existence of internal rehearsal processes for musical material.
- Mixed results have been brought for an *articulatory suppression effect* specific to verbal STM and not affecting musical STM. In a study using a backward recognition paradigm (same/different judgement between a first sequence and a second backward sequence, thus measuring WM rather than STM) for verbal, musical, and timbre material, Schulze et Tillmann (2013) showed that verbal articulatory suppression only affected performance for verbal WM, but not for the tone or timbre material. However, Schendel and Palmer (2007) conducted a recognition experiment on musically trained participants where they observed both a (sung) music and verbal suppression effect on both digits and tone sequences. Thus, the potential specificity of verbal articulatory suppression for verbal STM (that would suggest a separation between a sub-system for verbal material and sub-systems for other materials) remains unclear.

In addition to these effects, evidence has been brought that long-term knowledge of western musical structure has a beneficial impact on musical STM, equivalent to the beneficial effect of long-term lexical knowledge on verbal STM reported in the previous section. Schulze et al. (2012) and Albouy, Schulze, et al. (2013), using a recognition paradigm with simple tone sequences, found that non-musicians showed better performance for tonal sequences than atonal sequences (i.e., tonal material, that respects implicitly learnt western musical rules, is better memorized than atonal material). Additionally, Lévêque et al., (2022), using a recognition paradigm with more complex musical stimuli, showed similar beneficial effects of tonal structure knowledge on non-musicians' recognition performance. These results were interpreted as indicative of participants' implicit knowledge of tonal structure, acquired through exposure to music in everyday life, which could enhance their STM for structured material. These findings suggest that musical STM is influenced by long-term knowledge, similar to what is observed in verbal STM.

Overall, the observation of equivalent but specific behavioral effects for musical and verbal STM, in addition with the influence of long-term knowledge of implicit western musical structure on musical STM provide evidence that different mechanisms are at play, at least partly, for musical and verbal STM.

Studies comparing expert (musicians) or impaired (congenital amusics) participants to controls for music perception and memory has brought more knowledge about musical STM. Using recognition paradigms with verbal, non-verbal with contour, non-verbal without contour stimuli, for both auditory and visual materials, Talamini et al. (2022) showed that musicians had better STM performance for both auditory and visual contour stimuli, but not verbal stimuli. Thus, musical expertise is correlated to improved STM cross-modally when stimuli contain contour information. In a related vein, Chan (1998)

reported that musical training increases verbal WM performance. These results suggest common processes between short-term stores for auditory musical and verbal material (Chan et al., 1998) and between auditory musical and visual STM (Talamini et al., 2022). Conversely, using a recognition paradigm, Schulze et al. (2012) reported that when tonal and atonal melodies were compared in both groups, both musicians and non-musicians showed better performance for tonal sequences than atonal sequences, with a larger effect in musicians. For musicians, this effect was viewed as further evidence of the influence of long-term memory on STM performance, given their expertise in processing tonally structured material, similar to the long-term lexical knowledge influence on verbal STM, suggesting, as in non-musicians, separate processes between musical and verbal STM.

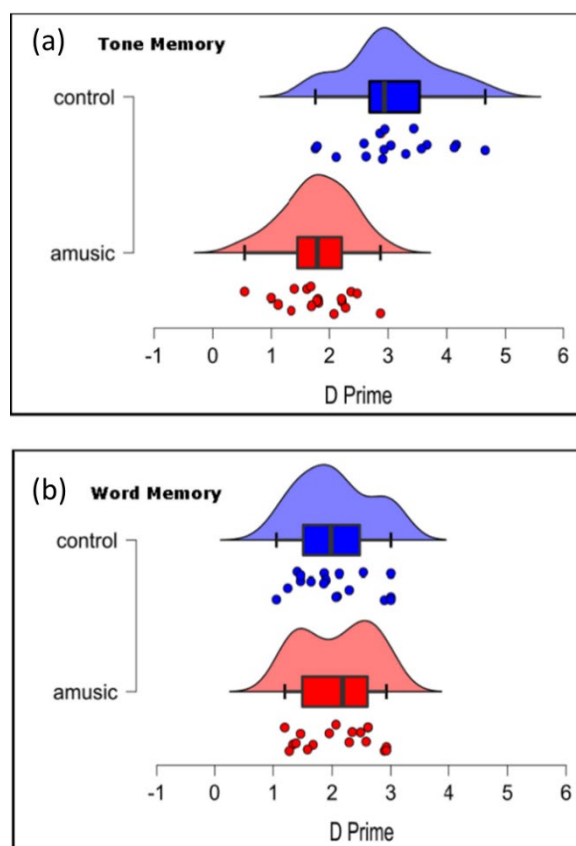


Figure 8: Performance of amusic ($n = 18$) and control ($n = 18$) groups in terms of sensitivity (d') in a delayed-matching-to-sample task using tones (a) and words (b).

A clear behavioral effect can be observed with impaired STM for tones for amusics and preserved STM for words. Figure from Tillmann et al. (2023).

Congenital amusia is a neuro-developmental disorder of music perception, production, and memory, notably pitch STM, with intact verbal and environmental sound processing (for reviews, see Peretz, 2016; Tillmann et al., 2015, 2023). The musical STM impairment observed in amusics is observed independently of pitch discrimination deficits (Tillmann et al., 2009) and cannot be attributed to a general STM impairment as amusic participants consistently show preserved STM for verbal material for both recognition tasks (Tillmann et al., 2009) and digit span tasks (Albouy, Schulze, et al., 2013; Williamson, McDonald, et al., 2010). Furthermore, congenital amusics' musical STM deficit becomes more pronounced when exposed to factors known to impair memory processing, including increasing the retention delay (Williamson, McDonald, et al., 2010), raising the memory load (Gosselin et al., 2009), and introducing interfering tones during the maintenance phase (Gosselin et al., 2009; Williamson & Stewart, 2010). These results provide further evidence supporting the notion of a distinct system for musical STM, as individuals with amusia show intact STM for verbal material (see Figure 8). However, despite their impairment, amusics appear to use, at least partly, the same system as controls for musical STM, as evidenced by the similar tonality effect observed in controls and amusics (Albouy, Schulze, et al., 2013; Lévêque et al., 2022).

In summary, the question of whether musical STM relies on a dedicated short-term store is still not definitively resolved. According to d'Esposito (2007), it seems counterintuitive that all temporarily stored information, including verbal, tonal, tactile, visual, and spatial, would require specialized dedicated buffers or systems. This perspective aligns more with functional WM

models that do not propose the existence of separate short-term stores for different types of materials (Barrouillet & Camos, 2007; Cowan, 2008). Progress have been achieved on this theoretical question through the investigation of these mechanisms using electrophysiology and neuroimaging (see section 1.2.2).

1.1.3.5 Development of auditory STM

Auditory STM play a key role in language and reading acquisition as holding auditory objects in memory is essential to learn how to understand and produce coherent communicative ensembles. Furthermore, auditory STM impairments are consistently reported in language-related learning disorders (Nithart et al., 2009), outlining its crucial role for a harmonious development. The majority of developmental studies on STM have primarily focused on verbal STM. Digit span, a common measure of verbal STM, has been found to increase with age until adulthood in various studies (Ahmed et al., 2022; Alloway et al., 2006; Bopp & Verhaeghen, 2005; Gathercole et al., 2004; Hale et al., 1997). Moreover, verbal STM performance has been linked to several other cognitive functions (Gathercole et al., 2001; Jones et al., 2020; Messer et al., 2015), highlighting its fundamental role in child development. However, our understanding of the developmental course of STM for other materials, such as music, remains limited. In the following sections, we will present current knowledge regarding the developmental trajectory of verbal and musical STM.

1.1.3.5.1 Development of auditory-verbal STM

As mentioned before, the majority of studies investigating verbal STM in children have used serial recall tasks involving the reproduction of short lists of verbal material like syllables, words, or digits in their original order. Based on these tests and research spanning several decades, there are established relationships between verbal STM and various domains in both typical and

atypical development including reading, vocabulary, reasoning, and problem-solving (Jones & Macken, 2018). Gathercole (1999) points out that “one of the most powerful factors influencing short-term memory capacity is age”. In a normative study using digit span task, Isaacs & Vargha-Khadem (1989) showed a gradual increase of performance in digit span in children from 7 to 15 years old. Chuah et al (1999) showed a linear increase of letter span between 5-6 year-olds and 8-9 year-olds and between 8-9 and 11-12 year-olds. Similarly, using digit, words and non-words span tasks, Gathercole et al. (2004) found a linear increase of performance in children from 4 to 14 years old, leveling off between 14 and 15 years. Using similar tests, Alloway et al (2006) found a linear increase of verbal STM between 4 to 10 year-olds, leveling-off between 10 and 11 years (see Figure 9). More recently, Ahmed et al (2022) used national representative data on a large sample ($n = 3652$) including digit span performance collected using an accelerated longitudinal design (i.e., testing all children with a wide age range at one point in time and testing them in two subsequent waves, 5 and 10 years later) in children from 3 to 19 years old. The authors demonstrated that children as young as 3 years old exhibited the ability to complete the digit span task. Furthermore, the most significant improvement in performance occurred during the age range of 3 to 10 years, after which there was a plateau until a short period of renewed growth between the ages of 13 to 16 years. These studies point out the importance of early and school-years childhood (3 to 10 years old) for the development of verbal STM. Several hypotheses have been posited about the sources of improvement of verbal STM throughout development.

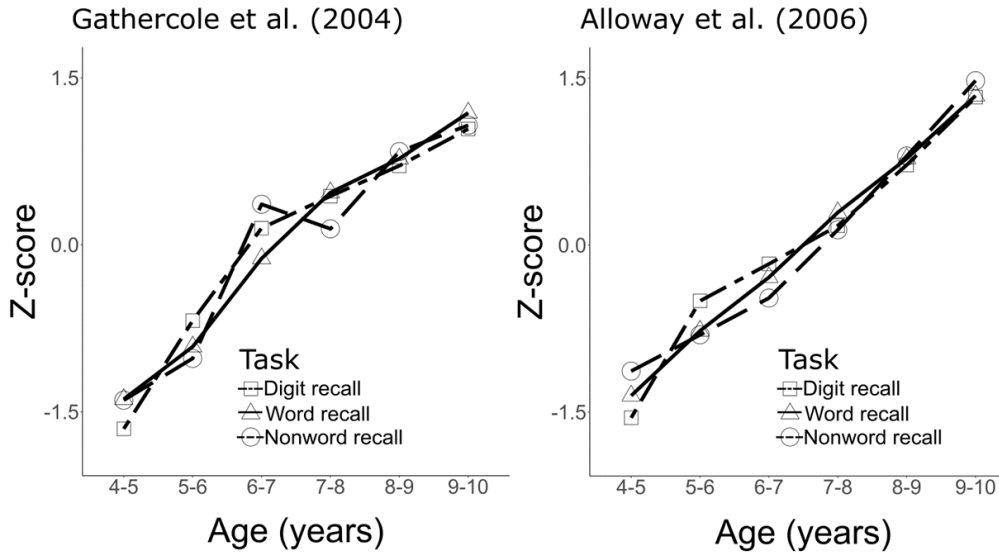


Figure 9: verbal STM developmental trajectory represented in z-scores of mean performance in verbal forward recall tasks with digits, words, and nonwords in children from 4 to 10 years of age from Gathercole et al. (2004) and Alloway et al. (2006). Figure from Ginzburg et al. (2022).

The substantial increase in memory capacity as children grow older is primarily attributed to the heightened rate of rehearsal, enabling them to retain larger amounts of verbal material in the phonological store (Gathercole, 1998). Younger children, below the age of seven, are not affected when visually presented pictures have lengthy articulatory labels or phonologically similar labels, suggesting that they do not recode pictures into their phonological forms (Hitch et al., 1989). Additionally, younger children do not overtly demonstrate rehearsal behaviors like lip movements or whispering, unlike older children (Flavell et al., 1966). However, when explicitly trained in overt and covert rehearsal strategies, 5-year-old children exhibit significant improvement in immediate serial recall for short words compared to long words (Johnston et al., 1987). This suggests that rehearsal strategies are present but not used spontaneously in children younger than 7 years old. Beyond the age of seven, an adult-like cumulative rehearsal strategy seem to emerge and maximize retention in the phonological store. It should be noted that studies that did not observe rehearsal processes in young children used visual stimuli that entails

recoding the items into their phonological form. Thus, the absence of spontaneous rehearsal could arise from difficulties in this recoding process. Using auditory stimuli could have been used to overcome the phonological recoding step and thus test rehearsal without an additional cognitive step that has been shown to be affected in younger children (Hitch et al., 1989). Despite the potential lack of spontaneous rehearsal in children below 7 years of age, STM span still increases during development, even before children learn to use rehearsal. One possible explanation is that older children can articulate items more rapidly during recall, leading to a reduction in the decay of memory items in the phonological store before output (Gathercole, 1998). This view is supported by the fact that articulation rate and memory span are not positively correlated in children below 7 years of age, while they are in older children (Gathercole et al., 1994; Gathercole & Adams, 1993; Henry, 1994; Keller & Cowan, 1994).

Another significant contributor to the observed increase in verbal STM during development is long-term linguistic knowledge. This knowledge may be associated with the semantic properties of verbal material (Allen & Hulme, 2006; Walker & Hulme, 1999), lexicality and frequency of occurrence of verbal items (Hulme et al., 1991, 1997), and correspondence between the phonological structure of the items and that of participants' native language (Gathercole et al., 1999; Jones et al., 2010). Additionally, it has been argued that domain-general processes (i.e., associative learning) might influence verbal STM as suggested by the facilitation effect of the familiarity of coarticulatory transitions on serial recall tests (Murray & Jones, 2002; Woodward et al., 2008) and prior exposure to inter-item transition probabilities (Botvinick, 2005; Botvinick & Bylsma, 2005; Majerus et al., 2012). Jones et al. (2020) went further by arguing that contrary to traditional views on age-related increases in verbal STM,

developmental changes in performance are not due to expanding dedicated verbal STM mechanisms but rather to the child's growing knowledge of the sequential structure of language acquired through linguistic experience.

Multiple factors contribute to the improvement in verbal STM performance during development and might collectively play a role in this phenomenon. STM for musical material has also been proved to play a role in the harmonious development of language and reading abilities, as evidenced by its impairment in language-related learning disorders. Thus, investigating the factors underlying its developmental trajectory seems essential to disentangle the shared and distinct processes between verbal and musical STM that might be impaired in learning-related learning disorders. In the following section, we will present the existing knowledge of STM development for musical material and propose that recognition paradigms are more suitable for comparing STM for both materials, as they do not require participants to reproduce the to-be-remembered items (which is impractical for items like tones).

1.1.3.5.2 Development of musical STM

To our knowledge, no specific study has focused on investigating the developmental trajectory of musical STM in children. However, a review of previous research reveals that pitch memory for single tones emerges as early as 6 months of age (Plantinga & Trainor, 2008). Research by Keller and Cowan (1994) found that the decay rate of STM for pitch reduces over time in children aged 4 to 12 years, as assessed through a two-tone comparison task. Specifically, they observed an increase in the persistence of memory for pitch between the ages of 6 to 7 years. Clark et al. (2018) investigated WM for musical material using a tone memory procedure where participants heard a sequence of non-musical tones and had to reproduce one of the tones with a 80-choice scale,

Auditory short-term memory

allowing to measure capacity (the number of tones held in working memory) and precision (how close to the queried target tone the reproduction was). They observed a linear improvement of WM capacity and precision from early childhood (around 6 years old) to pre-teenage years (around 13 years old Clark et al., 2018). Comparing it to a similar visual task, the authors observed that the developmental trajectory of WM capacity for tones resembled the developmental trajectory observed in visuo-spatial memory in children aged 6 to 13 years. These studies, though not numerous, suggest that STM/WM for musical material seem to undergo a developmental increase in childhood, as verbal STM does.

Despite the scarcity of developmental studies on musical STM, hypothesis can be put forward for the increase in capacity of musical STM (Clark et al., 2018). Keller et al. (1995) tested adults on a two-tone comparison task with and without rehearsing suppression and showed that performance was better when rehearsing was permitted . As for verbal STM, we can hypothesize that rehearsing of musical material is not spontaneously present before 7 years of age and that the absence of efficient rehearsing strategies could partly account for the developmental increase in musical STM. Moreover, musicians show better STM performance for visual or auditory materials with contour information than non-musicians suggesting that extensive exposure to materials with contour (a fundamental feature of musical information) might reinforce long-term knowledge of musical structures that in turn influences musical STM (Talamini et al., 2022). In the same way as the acquisition of linguistic knowledge and its influence on verbal STM, progressive acquisition of musical structure throughout development could also partly account for a developmental increase of musical STM.

Comparing musical and verbal STM is challenging when using standard recall paradigms, commonly employed for assessing verbal STM, as they often require production processes that are not easily adaptable for children. Some studies have attempted to compare musical and verbal STM in adults using mixed recall/reconstruction paradigms (Clark et al., 2018; Gorin et al., 2016, 2018; Williamson, Baddeley, et al., 2010), but they are closer to WM measures as they involve manipulation of the item held in memory while the participant reproduce a target item on a choice-scale. Consequently, when investigating musical STM or comparing musical and verbal STM, researchers typically employ recognition paradigms, with the classical delayed matching-to-sample task (DMST, Schulze et al., 2012; Schulze & Koelsch, 2012; Schulze & Tillmann, 2013; Talamini et al., 2022; Tillmann et al., 2009; for a review, see Caclin & Tillmann, 2018). In a DMST trial, the participant is presented with a first stimulus (S1), which may be an isolated sound or a sound sequence, and is required to memorize it. After a brief delay (usually a few seconds), a second stimulus (S2) is presented, and the participant's task is to determine whether S1 and S2 are identical or different. This task involves several processing steps, including encoding S1 into memory, retaining the information over the delay, and finally retrieving the memorized information to compare S1 and S2 (Figure 7b). As described in section 1.1.3.3, participant's performance and the degree of caution with which they respond can be addressed by using signal detection theory measures (d' and c , Macmillan & Creelman, 2005). Hence, the DMST appears to be a well-suited paradigm to assess the development of auditory STM for musical and verbal material (as an example, 11 out of the 18 studies comparing musical and verbal STM in adults reviewed in Caclin & Tillmann, 2018 used a DMST). The assessment of musical and verbal STM in school-age

children was the subject of the 2nd study of this PhD and is presented in section 3.2 (Ginzburg et al., 2022).

In this section on the development of auditory cognition, we reviewed current knowledge of three fundamental auditory processes, essential to the harmonious development of language and reading abilities: ASA, auditory selective attention, and auditory STM. Each of these processes undergo pivotal maturation steps during school-years up until adolescence and, although we approached each of these processes individually, they function in a complex interplay to make sense of our auditory environment. While the insights gained from the previously reviewed behavioral literature have greatly enhanced our comprehension of auditory cognition, certain questions remain unresolved or necessitate more in-depth exploration. To explore beyond behavioral outcomes, electrophysiological and neuroimaging methodologies have allowed us access to the physiological dynamics during the execution of these behaviors and brought much knowledge about auditory cognition and its development throughout childhood. The identification of objective neurophysiological markers of these processes are essential to increase our knowledge about the developing brain and pave the way for better prevention, care and remediation of learning disorders. The next section will thus be dedicated to present the cerebral networks subtending the auditory processes presented previously and a relatively new neuroimaging technique, functional near-infrared spectroscopy, that is particularly suited for the investigation of the auditory modality and for developmental research in typically developing and clinical populations.

1.2 Objective markers of auditory cognition: functional near-infrared spectroscopy

The aim of the present section is to introduce current knowledge about the investigation of auditory cognition with functional near infrared

spectroscopy (fNIRS). We will first guide the reader through current knowledge about the cortical networks subtending auditory processing, including speech processing. We will then delve into the neural correlates of verbal and musical STM, as it constitutes the core cognitive process explored with fNIRS in section 4 and 5. This neuroimaging technique being relatively new, the fNIRS technique will next be presented in detail, from its history to its data analysis. Finally, we will briefly overview the use of fNIRS in developmental neuroscience and its relevance for exploring auditory STM.

1.2.1 Auditory cognition: cerebral networks

1.2.1.1 Cortical networks of auditory processing

Auditory processing is supported by a large cerebral network involving, at the cortical level, various frontal, parietal and temporal regions. It has been suggested that, similar to the visual system (Rauschecker, 1998), auditory processing relies on distinct ventral and dorsal pathways, evidenced from both animal electrophysiological studies and human brain imaging research. Neurophysiological studies in non-human primates suggested ventral/dorsal routes supporting distinct “what” and “where” auditory processes (Kaas & Hackett, 1999). In this model, anterior temporal and inferior frontal areas process auditory “object” data, while posterior temporal, parietal, and frontal regions engage in spatial processing of stationary and moving sounds (Bendor & Wang, 2005; Poremba et al., 2003; Rauschecker & Tian, 2004; Tian & Rauschecker, 2004). In humans, various lesion, functional neuroimaging, and electrophysiological studies have aligned with this hypothesis (Alain et al., 2001; Anourova et al., 2001; Foster & Zatorre, 2010). However, the role of the dorsal pathway remains debated and numerous studies suggested that parietal regions are not only involved in auditory spatial processing but rather act as a

multimodal associative region receiving visual, auditory, and tactile information connected to frontal working memory areas and motor planning regions (Catani & de Schotten, 2008; Frey et al., 2008; see also Figure 10). Similarly, the specific role of the ventral stream for object identification has been challenged. Multiple brain regions, including the superior temporal gyrus (STG), parietal, frontal, para-hippocampal, insular, and occipital cortices, have been shown to be engaged in tasks like pitch or specific sound recognition (Alain et al., 2001; Maeder et al., 2001), suggesting that the ventral stream areas are not exclusively engaged in object recognition. Thus, assigning exclusive spatial and object functions to the dorsal and ventral streams seems insufficient (Rauschecker & Scott, 2009).

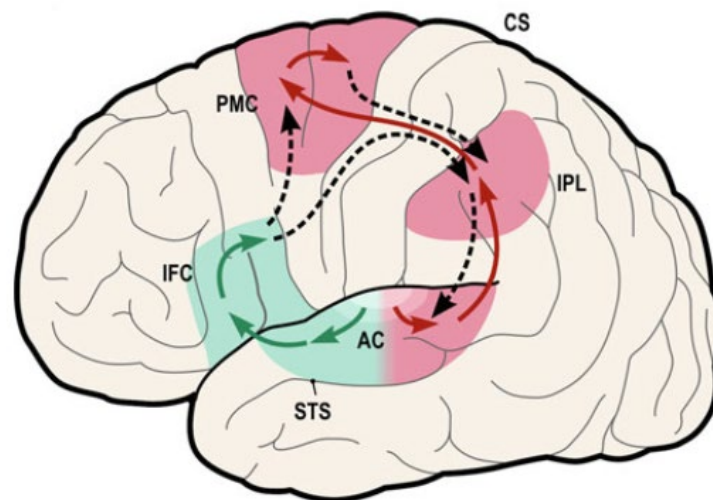


Figure 10: expanded model of dual auditory processing streams for speech (figure from Cohen et al., 2013; simplified from Rauschecker & Scott, 2009). The dorsal stream (red color) pivots around inferior/posterior parietal cortex, where a quick sketch of sensory event information is compared with an efference copy of motor plans (dashed lines). Thus, the dorsal stream plays a general role in sensorimotor integration and control. In clockwise fashion, starting out from auditory cortex, the processing loop performs as a forward model: Object information, such as vocalizations, speech or tones, is decoded in the anteroventral stream (green color) all the way to category-invariant inferior frontal cortex (IFC) and transformed into articulatory representations in the premotor cortex (PMC). Frontal activations are transmitted to the inferior parietal lobule (IPL), where they are compared with auditory and other sensory information. It is this fronto-parietal-sensory section that

expands the function of the dorsal stream beyond spatial processing only. AC, auditory cortex; STS, superior temporal sulcus; IFC, inferior frontal cortex; PMC, premotor cortex; IPL, inferior parietal lobule; CS, central sulcus

Based on cytoarchitectonic features, the superior temporal cortex in non-human primates comprises distinct functional areas, namely 'core', 'belt', and 'parabelt' regions, each exhibiting unique responses to auditory stimuli (Rauschecker, 2015). The lateral belt areas have shown heightened responsiveness to band-pass noise (BPN) over pure tones, displaying specificity for BPN center frequency and bandwidth. This suggests their role in decoding communication sounds that involve BPN bursts in various species, including humans (X. Wang, 2000). Neurons in the lateral belt also exhibit sensitivity to frequency modulation sweeps, indicating suitability for extracting communication sound features such as formant transitions in human speech. Human neuroimaging studies corroborate the organization of auditory cortex into core and belt areas, with the pure-tone responsive core region along Heschl's gyrus. This core region is surrounded by belt areas, activated primarily by BPN bursts, similar to non-human primates (Kuśmierek & Rauschecker, 2009). As in non-human primates, antero-lateral parts of the superior temporal cortex have been shown to respond to intelligible speech and speech-like sounds (Binder, 2000; Binder et al., 2004), while caudal belt and parabelt areas (projecting dorsally into the posterior parietal cortex) have been shown to preferentially engage in auditory spatial tasks (Rauschecker, 2015). These results highlight that the discrimination of behaviorally relevant auditory patterns, or “objects”, happen preferentially in the anterior ventral stream, contrarily to a long-standing view coming from stroke studies.

Indeed, for verbal material specifically, it had long been assumed that these recognition processes were located in the planum temporale, a posterior region of the superior temporal gyrus (i.e., “Wernicke area”). Contrary to these

notions primarily based on century-old stroke studies (Galaburda, 1978), more recent studies challenged such notions. A meta-analysis of over 100 human neuroimaging studies on phoneme, word and sentence recognition has identified an anterior location for the “Wernicke's area” rather than a posterior one, thus confirming the implication of the ventral pathway in speech processing (DeWitt & Rauschecker, 2012). This observation was later replicated by Yue et al (2019) who compared perception of speech versus non-speech (i.e. chords) and found more activation for speech material in the left superior temporal gyrus (see Figure 11).

However, speech processing is not confined to the superior temporal cortex. Perceptual invariance is crucial in the context of perceiving normal speech, where the “same” phoneme can exhibit significant acoustic variations due to coarticulation, yet remain identifiable as the same sound. This capacity to handle invariance is not limited to speech or auditory perception; it is a characteristic shared across various higher-order cortical perceptual systems (Rauschecker & Scott, 2009). The structural and functional organization of the anterior ventral streams in both visual and auditory systems offer insights into how the cerebral cortex addresses this challenge. Visual categories have been found to be formed in the lateral prefrontal cortex (PFC), which receives input from higher-level object representations (Freedman et al., 2001). In the auditory modality, research in macaque monkeys indicates the existence of neuron clusters within the ventrolateral PFC that encode complex calls, along with specific cells dedicated to individual semantic categories (Russ et al., 2008). In humans, the challenge of invariance in speech perception may be resolved in the inferior frontal cortex, or through interactions between the inferior frontal and anterior superior temporal cortex (see Figure 10; Rauschecker & Scott, 2009). The parietal and premotor cortices in the dorsal pathway also seem to

play a role in auditory-motor integration, control, and sequence processing (Rauschecker & Scott, 2009). Rauschecker & Scott (2009) proposed that these dorsal stream functions act as an interface between sensory and motor networks, performing a matching operation between predicted outcomes and actual events (see Figure 10 for a brief explanation).

While speech and music processing appear to engage comparable cortical networks (see section 1.2.2, there is a consistent observation of hemispheric specialization for each material (Zatorre et al., 2002). Under an acoustic-based account of auditory processing, the left hemispheric specialization for speech and the right hemispheric specialization for pitch-related aspects of music might stem from general-purpose mechanisms sensitive to the low-level acoustic features found in speech and music respectively (Poehpel, 2003; Zatorre et al., 2002). Supporting this view is the presence of spectrotemporal receptive fields in neurons all along the auditory pathway, identified both in humans and animal models (Massoudi et al., 2015; Schönwiesner & Zatorre, 2009). These neurons exhibit distinct sensitivity to spectral or temporal modulations in the auditory signal, offering a neurophysiologically plausible account for the neural decomposition of acoustic cues. Functional neuroimaging studies in humans reveal left hemisphere specialization in the auditory cortex for rapid temporal processing (Poehpel, 2003; Zatorre et al., 2002) and superior spectral resolution in the right hemisphere (Albouy et al., 2020; Zatorre & Belin, 2001). Moreover, an asymmetric sampling of time pattern has been observed in the left and right auditory cortices: the left auditory cortex appears to preferentially process auditory streams with a shorter temporal integration window (30 ms), while the right auditory cortex favors information sampling using a longer temporal integration window (200ms; Flinker et al., 2019).

Although the previously mentioned studies offer a plausible framework for a lateralized processing speech and music in the auditory cortex, some studies have proposed a category-based account of auditory processing restricted to associative cortical regions in the auditory ventral stream. Notably, an fMRI study showed that, regardless of semantic or basic acoustic characteristics, anterior regions of the secondary auditory cortices exhibited category-specific responses consisting in clusters selective to either musical instrument sounds or human speech (Leaver & Rauschecker, 2010). In contrast, regions of the secondary auditory cortex closer to the primary auditory cortex did not show such category-specific selectivity and rather responded to spectrotemporal acoustic attributes. Furthermore, no indications of lateralization were identified for the clusters sensitive to category, in contrast to the lateralization observed in clusters sensitive to spectrotemporal features.

Overall, a picture emerges of an anteriorly-directed hierarchical processing pathway, specialized in identifying and recognizing auditory patterns that are behaviorally relevant. Furthermore, neurophysiological spectrotemporal tuning to low-level acoustic cues appear to subtend a hemispheric specialization on speech and musical material in primary and secondary auditory cortices, while neural representations of increasingly complex categories appear to be encoded in the anterior regions of the secondary auditory cortex in a non-lateralized manner.

1.2.1.2 Cerebral correlates of speech processing in adverse conditions

Neuroimaging studies have revealed consistent patterns of brain activation when investigating speech recognition under various challenging conditions. Heightened speech intelligibility tends to elicit greater activity in the anterior and posterior regions of the superior temporal gyrus and sulcus

(i.e. the temporal region of the previously mentioned auditory ventral pathway). Under adverse conditions, opercular and cingulate areas are thought to underly speech recognition, forming a cingulo-opercular system (Eckert et al., 2016). Eckert et al. (2016) performed a meta-analysis of 10 neuroimaging studies on speech recognition in various challenging listening conditions including vocoded speech (Davis & Johnsrude, 2003; Hervais-Adelman et al., 2012; Wild et al., 2012), time-distorted speech (Adank & Devlin, 2010; Davis & Johnsrude, 2003), speech with unfamiliar accent (Adank et al., 2012), semantic ambiguity (Rodd et al., 2005), speech-shaped noise (Adank et al., 2012; Davis & Johnsrude, 2003; Golestani et al., 2013), and babble noise (Vaden et al., 2013; Wong et al., 2008). Their findings revealed that bilateral dorsal cingulate, inferior frontal, and anterior insula regions exhibited heightened activity when participants actively generated responses indicating stimulus recognition, in contrast to passive listening tasks. When examining closer the two studies involving babble noise (Vaden et al., 2013; Wong et al., 2008), increased activation in opercular and cingulate regions was associated with lower signal-to-noise (SNR) conditions as compared to higher SNR conditions. On the basis of these observations, Eckert et al. (2016) proposed that the cingulo-opercular system functions as an adaptive control mechanism to bolster speech recognition. This involves sharing performance outcomes (cingulate), enhancing arousal through autonomic function (anterior insula), augmenting response caution or inhibition (right inferior frontal cortex), and guiding controlled retrieval or response selection (left inferior frontal cortex).

Selective attention is a crucial process in speech recognition in adverse conditions by focusing on relevant information, suppress irrelevant information, and select a task-appropriate response (see section 1.1.2). A fronto-parietal system comprising the inferior frontal sulcus, the dorsolateral prefrontal cortex,

the intraparietal sulcus, and the inferior parietal lobule (Dosenbach et al., 2008) is thought to be involved in monitoring such functions during speech recognition in challenging situations (Eckert et al., 2016). Both younger and older adults exhibit an increased recruitment of the right dorsolateral prefrontal cortex (dlPFC) when speech quality is compromised compared to clear speech. Notably, in older adults, this fronto-parietal engagement increases throughout a task, suggesting an increased reliance on fronto-parietal control to counteract speech recognition deterioration arising from age-related auditory declines (Sharp et al., 2005). Interestingly, younger adults manifest heightened activation of the left dlPFC as speech intelligibility decreases, while older adults demonstrate the opposite pattern (Eckert et al., 2008). This might imply that older adults engage fronto-parietal control even under relatively manageable listening conditions, saving this control for more challenging tasks. In line with these findings, Wild et al. (2012) found that frontal region activity during speech recognition relies on the degree of attention directed toward stimuli, increasing when intelligibility is compromised by noise vocoding. Correspondingly, Zekveld et al. (2006) observed a non-linear response in frontal cortex activity while presenting sentences across various SNRs, with reduced activity evident in both the lowest and highest SNR conditions.

In summary, speech recognition in difficult listening conditions seems to rely on a cingulo-opercular network that remains engaged during task performance and adapts its activity based on task difficulty and success. This network appears to interact with a fronto-parietal attentional system to exert inhibitory control to suppress irrelevant responses and directing attention to relevant information. Just as frontal, parietal and cingular regions collaborate with temporal regions to optimize speech comprehension, auditory STM relies on a complex interplay of regions that facilitate the temporary storage of

auditory information. We will thus present in the next section the cerebral networks underlying auditory STM for musical and verbal material.

1.2.2 Auditory STM: cerebral networks

Through a synthesis of neuropsychological studies conducted on patients with brain lesions with neuroimaging and neurophysiological studies, neuroscience research has formulated a description of the cortical network that supports WM/STM processing. In the upcoming section, we will present current knowledge of the cerebral correlates of auditory STM for both verbal and musical material. We will then review studies that have concurrently explored auditory STM for both materials.

1.2.2.1 Cerebral correlates of verbal STM

By the early 1990s, the phonological loop emerged as the predominant model of verbal STM, thanks to its ability to explain various well-established behavioral effects in the field (Baddeley, 1992, also discussed in section 1.1.3). Functional neuroimaging offered a way to associate the phonological loop's components with distinct brain areas. Neuroimaging studies using positron emission tomography (PET) and fMRI explored the neural correlates of articulatory rehearsal processes and proposed that the left frontal inferior gyrus (IFG), along with the cerebellum and premotor regions, play a crucial role in verbal rehearsal (Awh et al., 1996; Gruber & Von Cramon, 2003; Paulesu et al., 1993; Ravizza et al., 2004). These findings imply potential similarities between the cortical networks underlying auditory verbal STM and speech perception and production (see section 1.2.1.1). However, results from neuropsychological studies indicate that localized temporoparietal lesions could lead to severe deficits in verbal STM while preserving speech comprehension and production (Baddeley et al., 1988; Shallice & Papagno, 2019; Shallice &

Warrington, 1970), suggesting that a separation exists between verbal STM processes and speech perception and production. A voxel-based lesion-symptom analysis involving 103 patients who underwent glioma resection surgery in either hemisphere confirmed these findings (Pisoni et al., 2019). The analysis revealed a connection between digit span scores and lesions in both the left supramarginal gyrus and superior-posterior temporal areas. In contrast, other verbal tasks including word comprehension primarily engaged areas that only partly overlapped with those involved in digit span tasks (e.g. middle and inferior temporal areas). These studies indicate that the neural underpinnings of auditory-verbal STM have only partial overlap with those supporting comprehension and production. While the left posterior-superior temporal cortex, associated with speech perception, contributes to both functions, the left supramarginal gyrus consistently and specifically underlies auditory verbal STM.

While lesion studies suggest a partial separation between speech perception and production regions and verbal STM regions, functional neuroimaging studies have yielded mixed results. Early PET studies, using various methodologies and tasks, initially associated the phonological store with the inferior or posterior part of the parietal lobe. However, Ravizza et al. (2011) found activation in the left superior temporal gyrus (defined a priori as a region of interest) instead of the expected parietal region during STM maintenance (compared to Jonides et al., 1997; Paulesu et al., 1993; Salmon et al., 1996). Unfortunately, many neuroimaging studies investigating verbal STM used visual rather than auditory stimuli for STM tasks (Braver et al., 1997; Langel et al., 2014; Lewis-Peacock et al., 2012; Narayanan et al., 2005; Rypma et al., 1999; Rypma & D'Esposito, 1999). Indeed, Lewis-Peacock et al. (2012) found that, using visual verbal stimuli, the occipital cortex played a prominent role

in distinguishing non-words from other stimuli during the retention period, creating uncertainty about whether participants were holding phonological, visual, or orthographic information. In contrast, Yue et al (2019) used auditory verbal material to investigate the separation from sensory and STM regions by comparing STM recognition tasks with perceptual tasks. During silent retention in STM tasks and using contrasts between two memory loads, they found activation in frontal and parietal regions, including the IFG, precentral gyrus, and supramarginal gyrus, but not temporal regions (see Figure 11 for a summary of their results). Similarly, Albouy et al. (2019) compared verbal STM with a control perception task using auditorily presented syllables. They found activations in left lateral frontal regions during maintenance in memory, including in the dlPFC and IFG, without involving temporal auditory regions.

In summary, the regions involved in maintaining speech in STM appear to be partly distinct from sensory areas. Frontal and posterior parietal regions seem particularly recruited for maintaining verbal information. The specificity of these regions for verbal STM remains a question, which will be developed in the subsequent sections focusing on the cortical networks underlying musical STM.

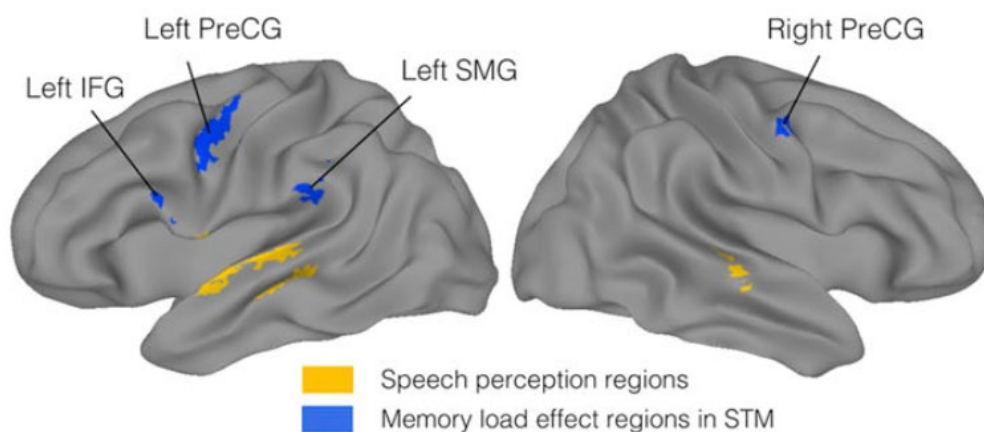


Figure 11: summary of Yue et al (2019)'s fMRI results. By comparing speech and non-speech perception, they observed significant activation only in superior temporal regions (in yellow). When localizing memory load effects in a STM task, they observed significant activations (in blue) in the left inferior frontal gyrus (IFG), the

left supramarginal gyrus (SMG) and bilateral precentral gyri (PreCG). Figure from Yue et al. (2019).

1.2.2.2 Cerebral correlates of musical STM

Compared to verbal material, the neural mechanisms underlying STM processing for musical material have received relatively less attention from both neuroimaging and neuropsychological investigations. In a lesion study, Zatorre and Samson (1991) found that auditory STM for musical material relies on the right auditory cortex. Nevertheless, this outcome was nuanced by acknowledging that a lesion in the auditory cortex could disrupt perceptual processing, likely leading to challenges in retaining an auditory trace over time. In addition to the involvement of auditory cortical areas, neuroimaging studies (Gaab et al., 2003; Griffiths et al., 1999; Holcomb, 1998; Zatorre et al., 1994) have revealed a cortical network remarkably resembling the one underlying the phonological loop described earlier. During active pitch retention, Zatorre et al. (1994) identified activations in the inferior frontal and insular cortex, planum temporale, and the supramarginal gyrus (SMG). This network was further observed by Gaab et al. (2003) during a pitch memory task, with activation in the SMG, the superior parietal lobule (SPL), planum temporale, premotor regions extending to the IFG, and cerebellum. Studies by Foster and colleagues emphasized the role of the intraparietal sulcus (IPS) in transforming pitch information during WM tasks, suggesting its contribution to information manipulation as well as maintenance processes (Foster et al., 2013; Foster & Zatorre, 2010). In a STM recognition task using tones, Linke et al. (2011) noted that while auditory sensory regions (like Heschl's gyrus) decoded tone frequency during the encoding stage, these same regions were suppressed in their representation during the maintenance stage. This suggests that, similar to verbal material, maintenance beyond sensory areas likely occurs in a distinct

representational format. Some studies (Albouy, Peretz, et al., 2019; Koelsch et al., 2009; Schulze et al., 2011) demonstrated that maintenance of musical information engaged lateral frontal, parietal, and subcortical regions without activating sensory regions. Collectively, these findings converge on the notion that STM for musical material involves interactions between frontal, parietal and posterior temporal regions, akin to observations made for other types of material (see Linke & Cusack, 2015 for STM for complex environmental sounds).

1.2.2.3 Comparison between verbal and musical STM

One approach to investigating whether there is a specialized brain subsystem for STM of different types of auditory materials is to compare how the brain handles auditory STM for verbal and musical materials (Albouy, Peretz, et al., 2019; Hickok et al., 2003; Koelsch et al., 2009; Schulze et al., 2011). In a study by Hickok et al. (2003), participants were asked to internally rehearse both verbal and tonal stimuli while undergoing fMRI scanning. The results revealed activation of the sylvian-parietal-temporal (Spt area), the IFG, and the left premotor regions during the internal rehearsal of both types of stimuli. Two other studies addressing the same question (Koelsch et al., 2009; Schulze et al., 2011) used recognition tasks and found very similar patterns of brain activation for both materials. Koelsch et al. (2009) observed that both verbal and musical processing during rehearsal activated the premotor cortex, anterior insula, SMG and IPS, planum temporale, IFG, and cerebellum. Schulze et al. (2011) also reported similarities in the brain activation patterns for the internal rehearsal of verbal and musical STM. In a more recent study, Albouy et al. (2019) examined STM for musical and verbal material using a recognition task. They compared the brain activations during the STM task with those of control perceptual tasks in both participants with amusia and matched control

participants. In control participants, while maintaining tonal and verbal information, activations were observed in left lateral frontal regions, including the dorsolateral prefrontal cortex (dlPFC) and the IFG, when comparing the memory task to a low-level perception task. Consistent with earlier research, both verbal and tonal STM tasks led to activations in brain regions commonly identified in previous studies focused on either verbal STM (Jonides et al., 1997; Paulesu et al., 1993; Salmon et al., 1996; Yue et al., 2019) or musical STM (Gaab et al., 2003; Griffiths et al., 1999; Hickok et al., 2003; Zatorre et al., 1994). Given the substantial overlap of neural resources supporting STM for both verbal and musical information, these studies concluded that there doesn't appear to be specific brain regions dedicated to the retention and rehearsal of musical material.

Another way to investigate the common and distinct cerebral correlates of verbal and musical STM is by examining potential neurophysiological distinctions between individuals with musical training and those without. Given that the processing of verbal material is a fundamental human ability generally acquired early in life, non-musicians can be considered specialized for speech processing, while musicians are specialized for both speech and musical material. Notably, there is compelling evidence indicating anatomical and functional variations between musically trained listeners and those without training in auditory processing (Baumann et al., 2008; Herholz & Zatorre, 2012; Schlaug, 2001, 2015; Stewart, 2008). Schulze et al. (2011) examined STM processing for verbal and musical information in extensively trained musicians and non-musicians. Brain regions such as the IFG, premotor cortex, insular cortex, cingulate gyrus, and IPL displayed stronger activation in non-musicians during verbal tasks compared to musical tasks. Conversely, these regions were more activated in musicians during musical tasks compared to non-musicians.

This suggests that the functional network used by non-musicians for verbal processing is also engaged by musicians for musical processing. However, it is worth noting that this analysis does not necessarily exclude the possibility that non-musicians engage the same brain regions but with less intensity. Furthermore, the authors proposed that musicians might use more refined sensory-motor codes for internally rehearsing tones compared to non-musicians, potentially due to the adaptability induced by musical training (Schulze et al., 2011; Schulze & Koelsch, 2012).

Another way to investigate differences between verbal and musical STM is by exploring anatomical and functional deficits in congenital amusia, a lifelong neurodevelopmental disorder of music perception and STM (for reviews, see Peretz, 2016; Tillmann et al., 2015, 2023). Since amusics present a specific impairment of STM for musical material but preserved verbal STM (Tillmann et al., 2009, 2016), they can be compared to controls to identify specific regions involved in musical STM. Several structural brain anomalies were revealed in congenital amusia including a reduction in white matter concentration in the right IFG (Albouy, Mattout, et al., 2013; Hyde, 2006), increased grey matter in the right IFG and the right auditory cortex (Hyde et al., 2007), a reduction in the connectivity between the right auditory and frontal cortices (Loui et al., 2009), and decreased whole-brain connectivity scores in congenital amusia (J. Wang et al., 2017; Zhao et al., 2016). Neuroimaging studies in congenital amusia have consistently identified abnormal activation and connectivity patterns in frontal and temporal areas when participants engage in perception tasks involving tones (Albouy, Caclin, et al., 2019; Albouy, Peretz, et al., 2019; Hyde et al., 2011). For musical STM specifically, Albouy, Mattout, et al. (2013) demonstrated functional abnormalities in frontal and temporal regions during the encoding, maintenance, and retrieval phases, providing evidence from MEG

recordings for anomalies in the fronto-temporal network involved in tone perception and STM. While the focus has been on music-related processes, it remained uncertain whether functional anomalies extend to verbal material. Albouy et al. (2019) conducted an fMRI study comparing STM tasks for both verbal and musical material against low-level perception tasks. Their findings revealed decreased connectivity between the right STG and right IFG in amusics during the encoding of musical material. During the maintenance phase, there were no differences between amusics and controls for verbal material. However, when it came to musical material, amusics exhibited decreased activation in the right frontal regions (IFG, dlPFC), right temporal regions, and left IFG, while showing increased activation in certain auditory regions compared to controls. These results suggest that during the maintenance of tonal information, amusics tend to rely more heavily on brain areas associated with auditory processing, reflecting an inefficient allocation of resources in STM networks for musical STM tasks.

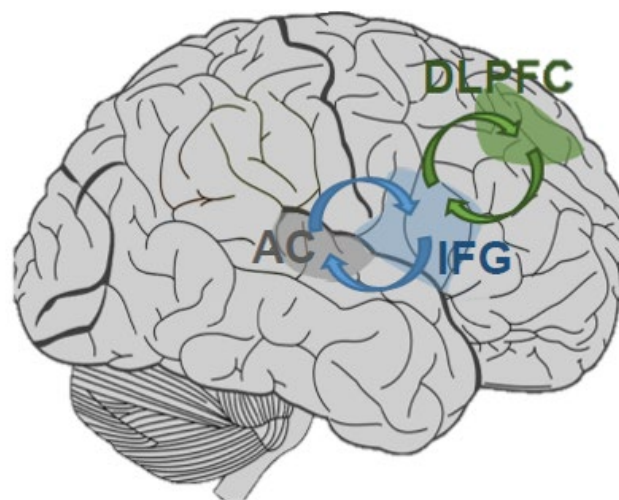


Figure 12: Dynamics of the cortical network of musical STM. Encoding a sequence of tones involves a loop between the auditory cortex (AC, in grey) and the Inferior Frontal Gyrus (IFG, in blue), particularly in the right hemisphere. During the maintenance of this sequence in memory, the dorsolateral prefrontal cortex (dlPFC,

Functional Near-Infrared Spectroscopy (fNIRS)

in green) is recruited. The recall of the information engages a network similar to the one involved during encoding (in blue). From Caclin (2021).

In summary, neuroimaging studies comparing verbal and musical STM support the notion of overlapping brain networks for both materials and evidence for distinct specialized brain networks dedicated to musical and verbal STM emerged from fMRI data when comparing populations with impairments (amusics) or expertise (musicians) to control groups (see Figure 12 for a model of the cortical networks involved in musical STM). However, the study of the auditory modality can be challenging when using fMRI, as particular considerations must be taken with the consequent noise it produces when recording participants. A relatively new neuroimaging technique, functional near-infrared spectroscopy, that we will present in the next section, has been showing from the past decades its suitability for the investigation of the auditory modality and for specific populations.

1.2.3 Functional Near-Infrared Spectroscopy (fNIRS)

1.2.3.1 History and instrumentation of fNIRS

fNIRS, an optical and noninvasive neuroimaging method, enables the monitoring of changes in concentrations of oxygenated (HbO) and deoxygenated (HbR) hemoglobin in brain tissue in response to neuronal activity. This is achieved by directing near-infrared (NIR) light (within the range of 650–950 nm) with a light source towards the head and recover it with a light detector placed on the scalp, a few centimeters away from the light source. The biological tissue's relative transparency to this range of NIR light allows it to penetrate and reach the brain tissue. The concept of cerebral tissue being penetrable by light dates back to observations by Richard Bright (1831), who noticed that the hydrocephalic brain of a patient became partially transparent in front of a candle. Much later, Jobsis (1977) formalized the idea

that both red and NIR light with longer wavelengths could pass through the scalp and skull to reach underlying tissues. Subsequently, a phase of exploration into muscle physiology and the clinical assessment of cerebral cortex pathophysiology followed (for a review on the history of fNIRS, see Ferrari & Quaresima, 2012). In the early 1990s, fNIRS recordings started, using single-channel devices, revealing fNIRS's ability to measure brain oxygenation and hemodynamic changes in response to functional activation tasks in adults and infants (Chance et al., 1993; Hoshi & Tamura, 1993a; Kato et al., 1993; Meek et al., 1998; Villringer et al., 1994). To harness the full potential of fNIRS, it became essential to employ multisite (or multichannel) measurements (Villringer & Chance, 1997). In the beginning, single-site devices were combined and employed simultaneously at multiple locations (Hoshi & Tamura, 1993b). Subsequently, the development of the first multichannel systems allowed for the monitoring of larger cerebral areas, facilitating the acquisition of topographic HbO and HbR change maps (Villringer & Chance, 1997).

Most of the presently commercially available fNIRS systems are based on continuous-wave (CW) technology, employing a continuously emitted NIR light, usually at two or three wavelengths (the present paragraph is mainly inspired from Scholkmann et al., 2014's review on fNIRS instrumentation and methodology). These systems measure the light attenuation of an emitted NIR light above the scalp due to tissue scattering and absorption by estimating the ratio of emitted to output light (Figure 13a and see section 1.2.3.2 for further explanation on signal measurement). By subtracting the first attenuation measure from subsequent attenuation measures, relative changes in attenuation are inferred and used to calculate changes in HbO and HbR concentrations. This approach assumes that changes in absorption are primarily dependent on oxygen-dependent hemoglobin chromophores, thereby excluding factors like

scattering, melanin, and water concentrations, which are unlikely to significantly vary during the measurement period. CW fNIRS devices offer insights into HbO and HbR concentration variations but are unable to determine absolute baseline concentrations, as they cannot distinguish and quantify absorption and scattering contributions. Apart from CW fNIRS technology, fNIRS instruments can also be classified into two other categories: time-domain (TD) and frequency-domain (FD) devices (see Figure 13b/c). Unlike CW systems, these methods enable the separation of light absorption and scattering contributions, allowing the determination of absolute HbO and HbR concentrations. FD devices employ intensity-modulated NIR light, while TD systems are more complex, employing a NIR light source with picosecond pulses and a rapid time-resolved detector to determine the time of flight of re-emerging photons (i.e., the time taken by light to travel through the tissue). These time-resolved systems offer lower time resolution, are costlier, and the time of flight tends to be noisier than intensity, rendering it less effective for detecting minor functional activations. In contrast, CW systems are relatively affordable, can be miniaturized and even wireless, and can be used in everyday situations. Due to these advantages and the emphasis on statistically detecting changes in brain activity rather than quantifying them absolutely in cognitive neuroscience, CW technology is the most prevalent and commercially available. As a result, the subsequent sections will exclusively focus on CW systems.

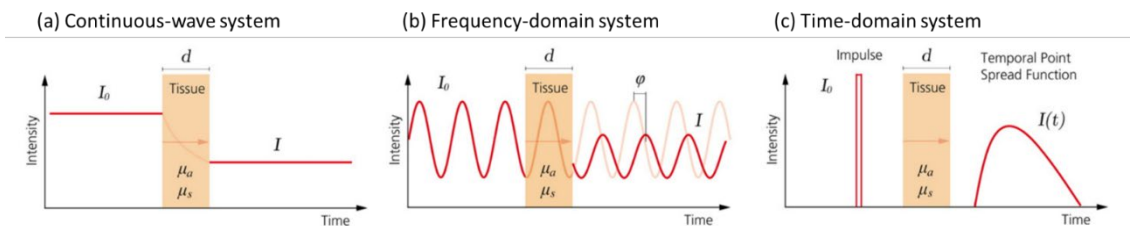


Figure 13: Illustration of the three different NIRS technologies. (a) continuous-wave technology involves emitting light with a constant intensity and subsequently monitoring alterations in the intensity of the transmitted light that travels through

the tissue. (b) frequency-domain technology modulates the intensity of emitted light and measures both the detected light's intensity and the phase shift, which corresponds to the time taken by the light to travel. (c) time-domain technology emits a brief light pulse into the tissue and measures the arrival times of photons emerging from the tissue. While this technology provides the most extensive information, it is also the most complex and costly approach. I_0 : incident light signal, I : transmitted light signal, d : thickness of the medium, μ_a : absorption coefficient, μ_s : scattering coefficient, φ : phase delay, and $I(t)$: temporal point spread function of the transmitted light signal. Figure adapted from Scholkmann et al (2014).

1.2.3.2 Physical principles

fNIRS measurements involve projecting NIR light onto the scalp. However, before it reaches the brain, the NIR light must navigate various layers like the scalp skin, skull, and cerebrospinal fluid, each with distinct optical properties. This interaction with human tissue is complex due to tissue anisotropy and inhomogeneity across layers. Yet, this can be simplified from recognizing that NIR light undergoes attenuation via absorption and scattering. Absorption refers to the conversion of photon energy into internal energy in the medium it travels through. This process depends on the molecular properties of the medium. Human tissues include numerous substances like water, lipids, hemoglobin, melanin, and cytochrome-c oxidase, exhibiting diverse absorption features at varying wavelengths (see Figure 14a). Notably, our body contains about 70% water, and its absorption in the NIR optical window is minimal, enabling the NIR light to traverse the tissue. Hemoglobin, a key absorber in the NIR window, exists in oxygenated (HbO) and deoxygenated (HbR) forms. HbO and HbR absorb light differently: HbO absorption is prominent at wavelengths $\lambda > 800$ nm (arterial blood, 98% saturation), while HbR absorption prevails at wavelengths $\lambda < 800$ nm (venous blood, 75% saturation). Besides absorption, NIR light scatters as it navigates through biological tissue. Scattering occurs more frequently than absorption and both result in light attenuation. Greater scattering entails longer photon paths, increasing its

probability of absorption. When light is projected into the head, it scatters, and penetrates several centimeters through the tissue. Consequently, positioning a light detector at a specific distance from the NIR source allows the collection of backscattered light (see Figure 14b), enabling the measurement of light attenuation. Given that NIR light absorption mainly arises from HbO and HbR, relative changes of the concentration of these chromophores during measurement can be estimated with a modified Beer-Lambert law that links the light attenuation at a specific wavelength to concentration changes in the corresponding chromophore (see Figure 15c, Pinti et al., 2020).

The association between a light source and a light detector is typically referred to as a “channel”. The region of tissue probed by the NIR light lies at the midpoint between the source and the light detector. The depth to which the light can penetrate (i.e., penetration depth) and then reach the detector is influenced by the distance between the light source and detector (i.e., longer distances result in deeper penetration). Numerous studies have investigated the spatial and depth sensitivity of fNIRS with respect to brain tissue, varying the source–detector separations and using Monte Carlo simulations (Brigadoi & Cooper, 2015). Increasing the distance between the light source and detector, which increases the proportion of detected photons that have traveled through deeper tissues, generally increases the sensitivity of the channel to brain activity (Strangman et al., 2013). However, this improved sensitivity to deeper tissues comes at the cost of a reduced signal-to-noise ratio (SNR) in the measured signal. This reduction in SNR occurs because fewer photons manage to reach the detector within a given time period without being absorbed (Brigadoi & Cooper, 2015). As a result, determining the appropriate source-detector separation requires finding a balance between depth sensitivity and SNR. Usual values that achieve this trade-off are source–detector separations of 30–35 mm

for adult subjects. While studies have established an optimal source-detector distance of approximately 20-25 mm for infants due to their thinner scalp and skull (Beauchamp et al., 2011; Brigadoi & Cooper, 2015; X. Fu & Richards, 2021), such investigations have focused solely on infants and not on school-aged children. The developmental growth of the brain-scalp distance during development occurs rapidly and in a non-linear manner, potentially reaching similar proportions to those of adults around 6 years of age (Beauchamp et al., 2011). However, there is currently a lack of data regarding the optimal source-detector distance in school-age children.

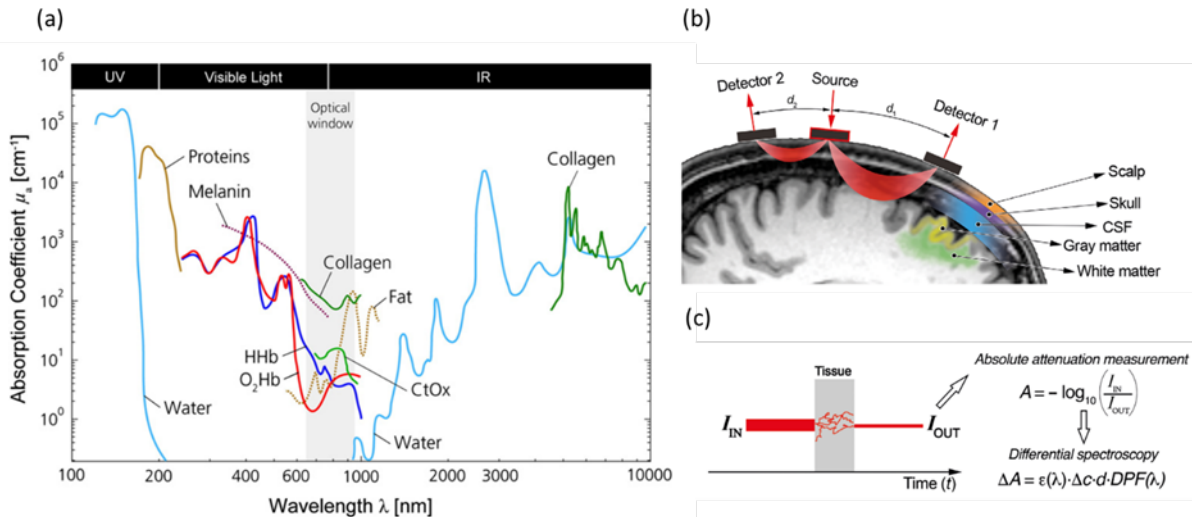


Figure 14: (a) Absorption spectra (log base) for different chromophores present in human tissue from 100 nm to 10,000 nm wavelengths (from Scholkmann et al., 2014). O₂Hb: oxygenated hemoglobin, HHb: deoxygenated hemoglobin, CtOx: cytochrome oxidase. (b) Schematic representation of the path (in red) followed by NIR photons from the light source to the detector through the different layers of the head. d_1 : deeper channel; d_2 : superficial channel (from Pinti et al., 2020). A channel is composed by a source–detector pair. (c) Concentration of HbO and HbR measures are obtained by subtracting all attenuation measurements by the first one (ΔA). Changes in concentration (Δc) are obtained by resolving the modified Beer-Lambert Law (differential spectroscopy equation). d represents the source–detector distance, ϵ is the extinction coefficient of the chromophore at a certain wavelength λ , DPF is the differential pathlength factor that indicates the increase in the photon path due to scattering (from Pinti et al., 2020).

1.2.3.3 Experimental design: constraints from hemodynamics

Neuronal activity triggers local arteriolar vasodilation, leading to an excess supply of blood termed functional hyperemia (Nippert et al., 2018). This enhanced blood flow is aimed at supplying the increased nutritional demands (i.e., glucose and oxygen) of active neurons. This results in a surplus of oxygen delivery with respect to its consumption in the activated brain area, causing an elevation in oxygenated hemoglobin HbO levels and a decrease in HbR levels (see Figure 15a). Referred to as a hemodynamic response, this phenomenon can be captured using fNIRS across various locations of the cerebral cortex. Typically, for a transient stimulus, the hemodynamic response peaks around 5-6 seconds after the onset of the stimulus and then gradually returns to its baseline state (Figure 15b), with a delay of approximately 16 seconds after the stimulus onset (Hoshi, 2016). The dynamics of this response, including the timing of its peak and the duration of its undershoot, can vary based on factors such as the brain region involved, the type of task, the age of the participants, and the time resolution of the measurement (Hoshi, 2016). These response dynamics play a significant role in shaping the experimental design.

In fNIRS research, different experimental designs can be used, but most studies tend to use either a block design or an event-related design. In a block design, participants are presented with a test stimulus for a fixed duration (usually from 5 to 30 seconds), alternating with a baseline period. The test stimulus needs to be long enough (typically 1-2 seconds) to generate a hemodynamic response (Wilcox & Biondi, 2015). In an ideal scenario, under controlled conditions, a hemodynamic response should be observable in a single trial. However, due to various factors like instrument noise, other physiological responses, issues with the coupling between the optodes and the scalp, and motion artifacts, the SNR often gets compromised. This leads to the need for a

larger number of trials for accurate data analysis. In block designs, the interpretation of data heavily depends on the nature of the baseline event (i.e., no stimuli or a control task) and the duration of the baseline interval. This interval needs to be long enough for the hemodynamic response to return to its original state, usually around 10 seconds (Hoshi, 2016). If participants are exposed to numerous trials, each followed by a baseline interval of the same duration, it can trigger anticipatory responses (Csibra et al., 2001). To address this, a jittered design is often used, where the length of the baseline interval is varied to make predicting the start of the test stimulus more challenging for the participant.

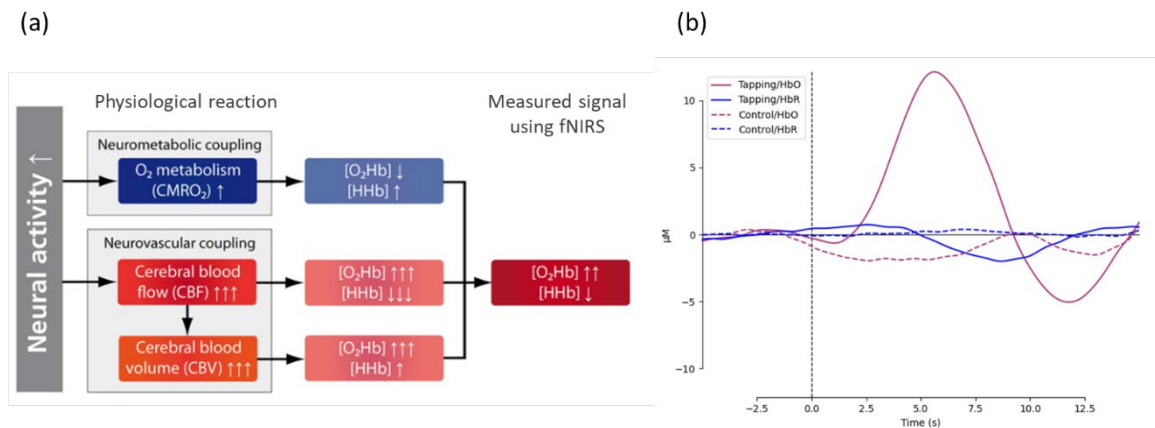


Figure 15: (a) Overview of the cerebral hemodynamic, oxygenation changes, and their impact on the measured fNIRS signals in case of an increased neural activity. CMRO₂: cerebral metabolic rate of oxygen, [O₂Hb]: oxyhemoglobin concentration, [HHb]: deoxyhemoglobin concentration. Figure from Scholkmann et al. (2014). (b) Example of measured changes in HbO (red lines) and HbR (blue lines) concentrations, averaged across 28 channels, during a 5-second finger tapping task (plain lines) and a control task (dotted lines). From pre-processing examples of the MNE-NIRS toolbox (<https://mne.tools/mne-nirs/stable/index.html>).

Given the resemblance of fNIRS signals to fMRI signals (hemodynamic responses), the event-related design and the general linear model (GLM) analysis widely used in fMRI studies have been deployed in fNIRS studies (Schroeter et al., 2004). In an event-related design, each task or stimulus is presented individually for a brief period. This approach allows for more

variability in the presentation of stimuli, rather than grouping them together as in block designs. Based on the assumption that hemodynamic responses add up linearly, statistical modeling using a GLM allows the fitting of explanatory variables (i.e. regressors) to each individual stimulus event (see section 1.2.3.5 for more details). This signal deconvolution process provides an estimation of how much each stimulus contributes to the overall hemodynamic response. The GLM allows greater statistical efficiency compared to simple averaging methods because the GLM considers the complete fNIRS time series and benefits from the higher sampling rate of fNIRS measurements (Pinti et al., 2020). Additionally, the need to wait for the hemodynamic response to return to baseline is less constraining in the event-related design, provided the jittering of interstimulus intervals to randomize the duration of overlaps between trials. This flexibility allows the use of more complex experimental designs (Petersen & Dubis, 2012). While theoretically more powerful, event-related designs include a decrease of signal-to-noise leading to less power than block designs of similar timing (Miezin et al., 2000), and more complex statistical treatment of the fNIRS signal.

1.2.3.4 Pre-processing of fNIRS data

After recording fNIRS data, four pre-processing steps are typically needed before making any statistical inference: (1) the conversion of raw signal to optical densities; (2) the conversion of optical densities to HbO/HbR concentration; (3) motion artifacts correction; (4) removal of physiological components unrelated to neuronal activity (Pinti et al., 2019; see also Figure 17).

The first step consists in transforming the raw signal (i.e., the attenuated NIR light measure) to relative changes of optical densities. For a single

recording channel and for one NIR wavelength, this is typically done by calculating the negative logarithm of the quotient of the division of the absolute intensity value of each data point by the mean of all intensity values of the channel during the recording session (Santosa et al., 2018). Note that since the relative change in optical densities is based on the average of all intensity values of the recording, it is important to preventively trim data that are not relevant to the task (breaks, instructions, or any events in which the participant is not engaged in the task).

Before continuing pre-processing steps, it is necessary to inspect signal quality and to exclude channels that have been jeopardized by environmental noise, poor coupling of optodes to the head, artifacts introduced by head motion, and optical interference by heavily pigmented hair (Hocke et al., 2018). One usually used technique is done by visual inspection of optical density data of the HbO-corresponding wavelength ($\sim 830\text{-}860$ nm) to check for the presence of cardiac pulsation that indicates that changes in optical density are coupled with physiological hemodynamic changes. However, there are several drawbacks of this method, namely its subjectivity, the fact that it does not ensure that HbR-corresponding wavelengths are of sufficient quality, and its time-consuming nature which makes it inefficient for large datasets. Consequently, several automated methods that allow the extraction of objective measures of signal quality have been proposed, including the scalp-coupling index (Pollonini et al., 2014) and the PHOEBE method (Placing Headgear Optodes Efficiently before Experimentation, Pollonini et al., 2016).

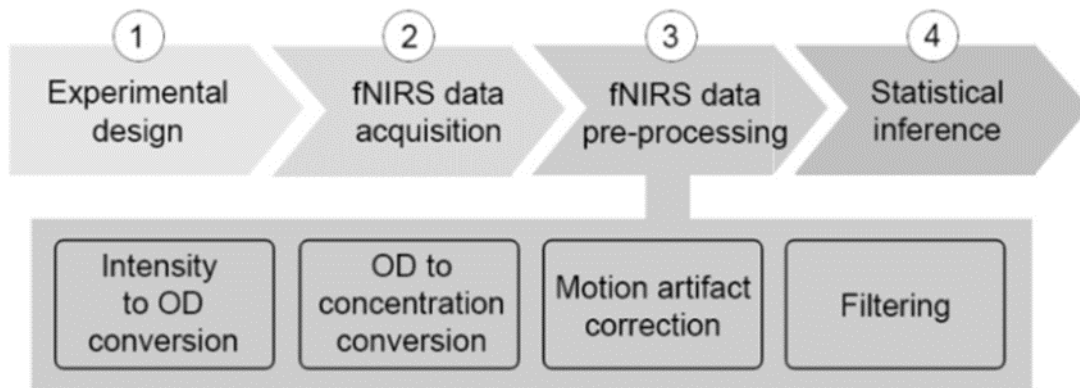


Figure 16 : typical neuroscience experiment pipeline using fNIRS. From Pinti et al. (2019).

The second step consists in transforming changes in optical densities to changes in HbO and HbR concentrations by resolving the modified Beer-Lambert equation (Delpy et al., 1988) that describes the changes in optical densities (i.e., loss of light intensity in tissue, also called attenuation) as a function of the chromophore concentrations (HbO or HbR), molar extinction coefficients, differential path length factor (DPF) that indicates the increase in the photon path due to scattering, and source–detector separation (see Figure 14c for the exact equation).

The third preprocessing step concerns motion artifacts. Due to the placement of fNIRS optodes on the participant's scalp, fNIRS is less susceptible to motion artifacts compared to stationary-sensor methods like fMRI. Nonetheless, motion artifacts can still pose challenges, especially in populations characterized by frequent and pronounced head movements, such as infants and children. A commonly employed approach to mitigate motion artifacts involves excluding trials where such artifacts are detected. However, this strategy is effective only when the number of detected motion artifacts is limited and the trial count is sufficiently high; otherwise, there is a risk of retaining too few trials, resulting in a noisy hemodynamic response (Brigadoi et al., 2014).

Moreover, this approach is mainly applicable in block design experiments, as event-related methodologies using GLM analysis must account for the entire recording time-course. In response to these challenges, various motion correction algorithms have been created, like the Temporal Derivative Distribution Repair (TDDR). This algorithm iteratively computes the weighted mean of the signal derivative and applies the corrected derivative to the data, resulting in the correction of baseline shifts and spike artifacts without the need for any parameters (Fishburn et al., 2019). Several other motion correction algorithms have been developed each with its own strengths and limitations (for comprehensive reviews, see Brigadoi et al., 2014; Cooper et al., 2012; Hu et al., 2015).

The fourth pre-processing step addresses the considerable variance due to hemodynamic fluctuations associated with physiological vasomotor regulations and breathing. These spontaneous components exhibit signals at distinct frequencies related to heart rate (~ 1 Hz), breathing rate (~ 0.3 Hz), Mayer waves (~ 0.1 Hz), and very low frequency (< 0.04 Hz) oscillations. A straightforward approach adopted by the scientific community is to employ digital filters to eliminate specific frequency bands in fNIRS signals. Based on a review of 110 fNIRS studies, Pinti et al. (2019) concluded that the most effective approach involves a band-pass Finite Impulse Response (FIR) filter. When selecting frequency cut-offs, they provide comprehensive guidelines that emphasize the importance of considering the frequency of task-specific stimulation. This ensures that the task-evoked response does not fall into frequencies excluded by the filtering process.

Unfortunately, certain spontaneous vasomotor components, notably Mayer waves (~ 0.1 Hz), often coincide with the frequency of task stimuli and cannot be effectively eliminated using conventional filtering techniques. To

address this issue, the fourth processing step can involve the application of principal or independent component analysis (P/ICA), which decomposes the signal and eliminates task-irrelevant components (Kohno et al., 2007; Y. Zhang et al., 2005). Alternatively, an approach to account for this spontaneous physiological signal is to directly incorporate corrections into the statistical model during the data analysis phase. The most common statistical method involves introducing regressors of no interest into a linear regression model during the statistical analysis step. These nuisance regressors can encompass external physiological measurements (e.g., finger pulse oximeter or a respiratory belt) or calculations of the systemic scalp response unrelated to neuronal activity derived from short-separation fNIRS measurements. Short-separation (SS) channels capture fNIRS signals for a 5-10 mm source-detector distance, theoretically recording only systemic signals, given the limited distance that prevents them from reaching cerebral tissue (Sato et al., 2016). In an extensive assessment of correction techniques applied to physiological signals, Santosa et al. (2020) concluded that incorporating SS channel information as a regressor of no interest within a GLM analysis yielded the most precise estimation of the underlying neural response, surpassing spatial and temporal filtering, regression, and component analysis. SS channels are being increasingly integrated into multichannel fNIRS montages and are now part of most commercially available hardware options.

It is crucial to emphasize that the pre-processing pipeline outlined above isn't composed of isolated stages. Each phase has a reciprocal influence on the others, and notably, they collectively impact the results of the statistical analyses and study findings. Currently, significant efforts are underway to establish a consensus regarding the variety of techniques and parameters

applicable along these steps (see Hocke et al., 2018, for a comprehensive review).

1.2.3.5 Data analysis

fNIRS data analysis techniques commonly fall into two main categories: averaging analysis and GLM analysis. In averaging analysis, fNIRS measurements are divided into segments centered around the stimulus onset and averaged together. On the other hand, GLM analysis involves fitting one or more model hemodynamic responses to the entire measured fNIRS signal.

Traditionally, fNIRS data has been examined using baseline-corrected averaging methods (Lloyd-Fox et al., 2010). This entails averaging the neural responses from blocks within a specific condition while applying a baseline correction. Subsequently, the processed fNIRS data is subjected to ANOVAs and/or paired sample t-tests conducted channel by channel to compare different experimental conditions (Filippetti et al., 2023). In this approach, the time course data is segmented into time intervals, such as baseline and experimental conditions. The mean concentrations of HbO and HbR during the baseline phase are subtracted from the trial window to calculate the mean changes in hemodynamic concentrations. These derived signals are then averaged across multiple trials for each channel and condition, followed by repeated measure analyses. Further post-hoc comparisons are made to control for potential false-positive activations using various methods for multiple comparisons. While the baseline-corrected averaging technique is relatively straightforward to apply to fNIRS data and avoids assumptions about the shape and timing of the hemodynamic response, it overlooks crucial temporal information (Tak & Ye, 2014). Additionally, it is only suitable for block-design experiments and cannot untangle the contributions of closely spaced events.

On the other hand, the GLM takes advantage of fNIRS data's high temporal resolution by considering the entire signal time course, making it a more robust analysis approach (Pinti et al., 2020). Although initially developed for fMRI data (Friston et al., 1994), the GLM has been adapted for optical data (Ye et al., 2009), capitalizing on the similarities between fMRI and NIRS designs and their reliance on the hemodynamic response. Two main approaches can be employed within GLM analysis. The most common approach involves modeling predetermined regressors, which are then convolved with a pre-established hemodynamic response function (HRF) and fitted to the data (Tak & Ye, 2014). However, the drawback of this HRF-based GLM method is its assumption of a fixed HRF shape. This assumption may not hold consistently within and between subjects, especially in newborns and young infants (Filippetti et al., 2023). An alternative method in GLM analysis is the use of a Finite Impulse Response (FIR) deconvolution model. Adopted from the fMRI domain (Henson & Friston, 2007), this approach does not assume a predefined hemodynamic response shape. Instead, a sequence of impulses occurring after the stimulus onset is employed as regressors to model the neural response (Luke et al., 2021). While computationally more demanding, this method enables the reconstruction of the hemodynamic response to a particular condition around its onset, without presuming its shape (as with the averaging approach, see Figure 17). Additionally, it retains the key benefit of the GLM approach – the incorporation of the entire signal (see section 4 and 5 for the use of this approach on fNIRS data). Irrespective of the chosen GLM approach, once beta coefficients are obtained for each regressor (for detailed mathematical models, see Huppert, 2016), second-level statistical analysis can be applied to these coefficients to compare conditions or groups (Santosa et al., 2018).

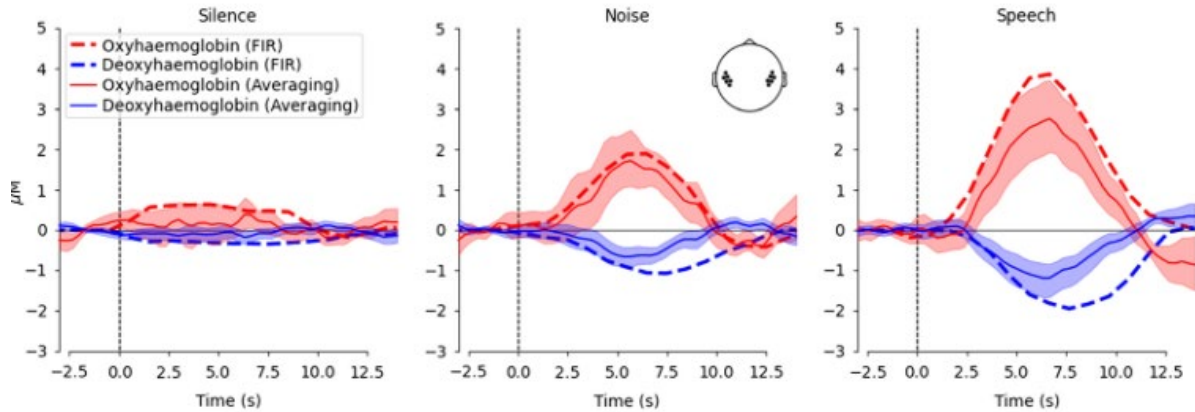


Figure 17 : morphology of auditory fNIRS responses over the superior temporal gyrus during passive listening of silence, noise and speech. Changes of HbO (in red) and HbR (in blue) concentrations are shown after using an averaging (plain lines) approach or a finite impulse response (FIR) deconvolution model (dashed lines). Note that for a block design, both approaches yield similar results. Figure from Luke et al. (2021).

Additionally, recent advances have been made in applying multivariate pattern analysis (MVPA) on fNIRS data. These multivariate techniques can differentiate task-related brain activity across two or more experimental conditions. In contrast to conventional univariate tests, MVPA holds the potential to provide significant details into aspects of the hemodynamic signal that might be discarded when using traditional univariate tests (Emberson et al., 2017).

1.2.3.6 Limitations of fNIRS

While fNIRS offers distinct advantages, it does come with limitations. Its temporal resolution is lower compared to EEG or MEG, and its spatial resolution is lower than fMRI. As a result, fNIRS is limited in its ability to probe cerebral hemodynamic responses within deeper brain regions. Its scope is mainly restricted to measuring superficial cortical activity, with the typical depth sensitivity of fNIRS systems reaching only around 1.5 cm. The minimal separation required between light sources and detectors for penetrating brain

tissue poses constraints on spatial resolution and the number of available recording optodes. Additionally, individuals with high hair or skin pigmentation may drastically increase NIR light absorption, reducing SNR and complicating measurements in certain populations. Given its relatively recent emergence, there remains ongoing debates within the fNIRS community regarding the consensus on preprocessing and analyzing data. Aspects such as acquisition techniques, signal processing methods, statistical analyses, and data interpretation lack standardized procedures, which can be challenging for newcomers to the field. Despite these challenges, effort is being made within the fNIRS community to share information and provide useful knowledge, thus creating a stimulating environment conducive to innovation and progress about its application from developmental to clinical science (including in the agronomic field for identifying stored-grain insects, Dowell et al., 1999).

1.2.4 fNIRS in developmental science

fNIRS offers a well-suited method for investigating the cognitive processes of children and infants (Aslin, 2012; Chakravarti et al., 2008; Cristia et al., 2013; Ferreri et al., 2014; Lloyd-Fox et al., 2010; Soltanlou et al., 2018; Wilcox & Biondi, 2015). It has been proposed that fNIRS serves as a bridge between current knowledge of cortical activity in developing brains and our understanding of adult human brain function (Lloyd-Fox et al., 2010). The non-invasiveness, absence of side effects, minimal need for immobilization, reduced sensitivity to motion artifacts, and silent nature of fNIRS addresses many obstacles encountered in developmental neuroscience. A great deal of the developmental science work using fNIRS has focused on the cortical aspects of language development (Wilcox & Biondi, 2015). For instance, Tellis and Tellis (2016) explored brain activation changes in children and adults during silent reading, reading out loud, and free speech. They observed the highest activation

during free speech in bilateral frontal regions in participants from 11 to 52 years old. A study by Tando et al. (2014) identified increased activation in the frontopolar area from ages 6 to 18 years during a verbal fluency task. Additionally, Kawakubo et al. (2011) observed developmental changes from 5 to 37 years in the frontopolar region during a letter fluency task. Examining pre-literate children (average age = 4.2 years) over a one-year span for reading abilities, Jasinska et al. (2021) found that functional connectivity between frontal (IFG) and temporal (STG) regions correlated with developmental changes in reading ability. Moreover, the exploration of brain mechanisms in bilingual children using fNIRS has generated significant interest within the field of language research. Researchers have found functional differences in temporal and inferior frontal regions when comparing monolingual and bilingual children using reading tasks (Jasińska et al., 2017; Jasińska & Petitto, 2014), verbal fluency tasks (Mücke et al., 2018) or sentence processing tasks (Jasińska & Petitto, 2013). Interestingly a large-sample study has been conducted on 484 elementary-school children (aged 6 to 10) who engaged in word repetition tasks in both their native language and second language. Researchers acquired fNIRS recording on-site with a fNIRS vehicle, showcasing the portability and adaptability of fNIRS devices (Sugiura et al., 2011).

In addition to its role in understanding language development, fNIRS has been employed to investigate neural mechanisms in children with various developmental disorders. Autism Spectrum Disorder (ASD) has been a focus of investigation, with comparisons between ASD and typically developing (TD) children revealing functional distinctions in various cognitive processes (for a review, see F. Zhang & Roeyers, 2019). For example, infants at elevated risk of ASD have demonstrated diminished activation in temporal and frontal cortices during social perception tasks (Braukmann et al., 2018), and abnormal

functional activity related to face processing has been identified in the ASD brain (Fox et al., 2013). Language processing studies have indicated altered inter-hemispheric functional connectivity between temporal cortices and reduced connectivity between temporal cortices and inferior frontal regions in children aged 8 to 11 with ASD compared to TD children (Zhu et al., 2014). Regarding executive functions, fNIRS studies have uncovered inhibitory dysfunction in individuals with ASD, characterized by reduced activation in the right prefrontal cortex during a go/no-go task (Xiao et al., 2012) and an auditory attention task (Funabiki et al., 2012). Moreover, fNIRS has been extensively applied to the exploration of attention deficit hyperactivity disorder (ADHD). Findings have indicated that ADHD-related symptoms are associated with decreased cortical activity in regions like the inferior prefrontal gyrus, middle prefrontal gyrus, supra-marginal gyrus, and angular gyrus, during tasks involving prosody processing (Köchel et al., 2015), emotion recognition (Ichikawa et al., 2014), suppression of irrelevant stimuli (Inoue et al., 2012; Jourdan Moser et al., 2009; Monden et al., 2015), olfactory recognition (Schecklmann et al., 2011), and working memory (Schecklmann et al., 2010; Tsujimoto et al., 2013). fNIRS has also brought significant insights into other neurodevelopmental disorders like stuttering (Chang, 2014) and Down syndrome (Xu et al., 2020).

Surprisingly, the use of fNIRS for exploring language-related learning disorders has been relatively limited. A systematic review by Butler et al. (2020) highlighted only three empirical fNIRS studies investigating the neural correlates of language processes in children with language impairments. Fu et al. (2016) tested children with developmental language disorder (DLD), using a non-parametric statistical approach to evaluate language comprehension. Their findings revealed disparities in HbO trends within bilateral IFG and left

inferior posterior parietal regions between DLD and TD children along with differences in HbR trends in the right inferior posterior parietal cortex and left temporo-parietal junction. Cutini et al. (2016) used fNIRS to assess the neural correlates of amplitude-modulated white noise stimulation at 2 Hz and 40 Hz in children with dyslexia and TD children. Their study unveiled a significant difference in the hemodynamic response to 2 Hz stimulation in the right supramarginal gyrus between dyslexic and TD children, while no difference was observed for the 40 Hz stimulation. In the context of the temporal sampling framework, the authors suggested that dyslexics might be impaired in the phonetic rate processing of language, corresponding to the 2 Hz stimulation. Likewise, Song et al. (2013) identified lower HbO levels in the left dlPFC and IFG during a consonant-vowel task in dyslexics compared to controls. More recently, Hancock et al. (2023), in a WM task with TD and DLD children, reported an increase in HbO changes in response to increasing memory load in the left dlPFC, while bilateral inferior parietal lobules showed a decrease in HbO with increasing memory load within the TD group but not in the DLD group.

In summary, fNIRS proves highly suitable for investigating various cognitive processes including language and language-related higher-order functions such as STM or WM in typical and atypical development. In the case of language disorders, even though this area remains relatively unexplored with fNIRS, it holds a promising potential to provide valuable insights into the complex interplay between brain function and developmental trajectories.

1.2.5 Exploring auditory STM with fNIRS

fNIRS studies that aimed at exploring functions in which auditory material is maintained have mostly done so by manipulating memory load using

verbal n-back tasks (see Figure 7c). In this task, participants are presented with a series of stimuli (e.g., letters, digits) and are required to indicate when the current item matches the one presented n trials earlier. Most neuroimaging studies (e.g., fMRI) used verbal n-back tasks to manipulate memory load and found that activity consistently increased with memory load in lateral prefrontal regions (for reviews, see Owen et al., 2005; Rottschy et al., 2012). Some fNIRS studies have used verbal n-back task to manipulate memory load (for a review of these studies, see section 4.2, pages 217-219) and found similar results as in fMRI studies. However, the n-back task entails not only the maintenance of information in memory but also the updating of such information with each upcoming stimulus, thus tackling WM processes instead of purely STM. With the aim to target STM specifically, behavioral and neuroimaging studies have successfully shown that memory load manipulation could be achieved with DMSTs (behaviorally with words, pitch and timbre sequences, e.g., Schulze et al., 2012; Schulze & Tillmann, 2013; using MEG with tone sequences, Grimault et al., 2014). To our knowledge, only one methodological fNIRS study has explored auditory STM using a DMST paradigm with auditorily or visually presented syllables (Yamazaki et al., 2020). After listening to (or watching at) a 9-syllable sequence, participants had to maintain the information during a 9-second retention delay and compare it to a second 9-syllable sequence that could be either identical or different by one syllable. Within a large array of recording channels over the left frontal and temporal areas, significant activation during the encoding and maintenance phase was observed in the auditory modality in the IFG and dlPFC respectively, along with other premotor and temporal areas, in keeping with fMRI studies (see section 1.2.2). The investigation of auditory STM and

memory load manipulation for both verbal and musical material will be the subject of the study presented in section 4.2 of this PhD.

To the best of our knowledge, no fNIRS study investigated auditory STM in children. We found only one study that manipulated memory load with an *auditory* n-back task in DLD and TD children (Hancock et al., 2023, see section 1.2.4). Other fNIRS studies that investigated memory load in verbal WM have done so by using *visual*-verbal n-back tasks. Han et al. (2022) compared high-functioning ASD children with TD children with a digit n-back task with three memory loads (0-back, 1-back, and 2-back). The study revealed that when comparing the 0-back and 1-back loads, there was a noticeable rise in functional connectivity in the right medial PFC within the TD group. In contrast, the group with ASD exhibited a declining trend in right medial frontal intrahemispheric connectivity with increasing load. Yeung et al. (2019) conducted a similar study on adolescents with ASD using consonants for a verbal n-back task with three memory loads (0-back, 1-back, and 2-back). They found that adolescents with ASD showed more right-lateralized prefrontal activation in response to WM load (i.e., 2-back > 0-back) as compared to TD adolescents. These studies show that manipulating memory load in children with developmental disorder can bring valuable insights into the functional correlates of the atypical development. Preliminary results of the exploration of verbal and musical STM in children with fNIRS will be presented in section 5.2 of the present PhD.

In conclusion to this theoretical background section, we showed that the harmonious development of language and reading depends on several auditory cognitive processes that act together in a complex interplay. These processes undergo crucial maturation steps during a child school-years and later in adolescence. The factors underlying their development are numerous and more

investigation is needed to uncover and precisely characterize them. While electrophysiological and neuroimaging studies have significantly expanded our knowledge about the cortical underpinnings of auditory cognitive processes, efforts are still being made to develop new hardware and analysis tools to unveil objective markers of such processes. In this context, the objectives of the present PhD work are presented in the next section.

Objectives

The overall objective of this PhD was to explore auditory cognition processes (1) behaviorally in children and (2) neurophysiologically in adults and children (including with learning disorders) with a novel neuroimaging tool. Studies presented in section 2 and 3 describe the developmental trajectory of two key processes of auditory cognition in children from 5 to 10 years old: speech perception in speech noise and auditory STM for verbal and musical material. Studies in sections 4 and 5 describe the prefrontal involvement in auditory STM processing for musical and verbal material in adults and children using functional near infrared spectroscopy (fNIRS).

Speech in babble-noise perception is a crucial auditory process that children encounter in most social and learning environments and that does not seem to be fully mature until adolescence. In the study presented in section 2 we tested 100 school-aged children with a newly developed closed-set speech perception in babble-noise test combining four levels of auditory and phonological difficulties. As children with language-related learning disorders consistently show impairment in speech perception in noise, our aim was to characterize its developmental trajectory with a short, child-friendly, and sensitive screening test. These investigations and their associated findings led to the submission of an article, currently under review.

Auditory STM plays a central role in most communicative situations and is essential to the optimal acquisition of language as evidenced by the consistent impairment of auditory STM for musical and verbal material in children with learning disorders. In the study presented in section 3, we assessed the developmental trajectory of musical and verbal STM by mean of a delayed matching to sample task (DMST) in the same children as in section 2. To our

knowledge, this was the first study to compare STM in children for musical and verbal material using the same task for both materials. The aim of this comparison was to unveil domain-general and domain-specific processes related to the development of STM for the two materials. In addition, we explored common processes between auditory STM and speech perception in babble-noise. This study led to a paper published in *Developmental Science* (Ginzburg et al., 2022).

fNIRS is a particularly well-suited technique for exploring the auditory modality and is thus a good candidate for the exploration of the neurophysiological dynamics of auditory STM in children with learning disorders. As to date, a relatively small number of studies explored these questions, we first conducted a study in healthy adults and a second one in typically developing (TD) children and children with learning disorders.

The study in section 4 aimed at using fNIRS to record prefrontal areas to identify objective markers of auditory STM. The first experiment acted as a proof of concept in healthy adult participants as to our knowledge, no fNIRS study explored auditory STM with a DMST for musical and verbal material. We thus aimed, in the first experiment, at replicating the findings of an fMRI experiment (Albouy, Peretz, et al., 2019) that showed the involvement of lateral prefrontal regions during the maintenance of musical and verbal material in STM. In a second experiment, we manipulated memory load in musical and verbal STM to identify the recruitment of lateral prefrontal regions when memory load increases, and explore their potential differences in dynamics for musical and verbal material. This study led to the submission of an article, currently under review.

Based on the successful experiments that we conducted on adults with fNIRS, we explored the memory load effects observed in adults in school-aged children with language-related learning disorders and TD children. The recruitment and testing of children being still in progress, we will present preliminary results on 11 TD children in section 5.

2 Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

Article under review

2.1 General Introduction

Speech recognition in babble noise is a challenging situation that children encounter throughout their entire development in numerous learning environments, notably their classroom where they have to focus on a teacher's voice while ignoring multiple irrelevant sound sources (e.g., background whispers or conversations, clatters of objects falling on the classroom floor, etc.). While healthy normal-hearing adults perform this task in a seemingly effortless way, this ability does not reach adult-like performance in children up until adolescence. In the present study, we tested speech in babble-noise recognition in school-aged children, a critical maturation period for speech-in-noise perception, with a closed-set, child-friendly, easy-to-use screening test, combining four levels of phonological and signal-to-noise difficulties, and we compared their performance with a group of young normal-hearing adults. The present report is the manuscript of an article submitted for publication.

2.2 Article

Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

J. Ginzburg¹, L. Fornoni¹, P.E. Aguera¹, C. Pierre¹, A. Caclin¹, A. Moulin¹

¹ Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France

Running title: speech-in-babble-noise development

Acknowledgments: We thank all the children who took part in the study, along with their parents, and their welcoming teachers. The data necessary to reproduce the analyses presented here are not publicly accessible due to privacy or ethical restrictions. The data are available on request from the corresponding author. The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible at the following URL: https://github.com/jeremieginzburg/cd_analysis_sample_code. The materials necessary to attempt to replicate the findings presented here are not publicly accessible. The analyses presented here were not preregistered.

Funding: This work was conducted within the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) and of the LabEx Cortex (“Construction, Cognitive Function and Rehabilitation of the Cortex”, ANR-11-LABX-0042) of Université de Lyon, within the program “Investissements d’avenir” (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). This work was funded by a Pack Ambition Recherche (COGAUDYS project) from the Auvergne-Rhône-Alpes Region to AC and AM.

Abstract

Speech-in-noise perception is consistently reported to be impaired in learning disorders, which stresses the importance of documenting its developmental

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

course in young children. In this cross-sectional study, ninety children (41 females, 5.5-11.6 years-old) and nineteen normal-hearing adults (15 females, 20-30 years-old), were tested with a newly developed closed-set speech perception in babble-noise test, combining four levels of phonological and auditory perception difficulties. Results showed that speech-in-babble-noise perception takes a definite maturation step around 7 years-of-age ($d = 1.17$, grade effect) and is not mature at 10 years-of-age when compared to young adults ($d = .94$, group effect). Developmental trajectories of both accuracy and response times were evaluated, with influences of psycholinguistic factors, to foster the development of adequate screening tests.

The ability to make perceptual sense of a complex sound environment (Auditory scene analysis, Bregman, 1994) is a key component of auditory cognition and a major element for harmonious cognitive development in children. If hearing screening is now well established in many countries, from birth and at regular interval during primary school years in children, this screening is aimed at detecting peripheral hearing loss, which is of paramount importance (Skarżyński & Piotrowska, 2012), but such screening only occasionally involves speech audiometry (in silence) and only exceptionally, speech perception in noise. However, an estimated 7% of children (Bamiou et al., 2001; W. J. Wilson & Arnott, 2013) show specific difficulties in listening in noise with normal audiograms, which is one of the characteristics of central auditory processing disorders. Speech-in-noise perception, along with other key auditory cognition processes (e.g., auditory short-term memory, STM), are consistently reported to be impaired in learning disorders (Bradlow et al., 2003; Majerus & Cowan, 2016; Plaza et al., 2002; Ziegler et al., 2005, 2009). Specific listening difficulties in noisy environments in the absence of any peripheral hearing loss are rarely isolated (Moore, 2018). They are in fact associated with other deficits (such as attention-deficit and hyperactivity disorders, speech and language impairments, and impaired manual dexterity) in 90% of the cases according to Ahmmed (2021). This fact stresses the need for systematically screening such listening difficulties specific to noisy environments (Skarzynski et al., 2015), as they are often associated with learning disorders that will sometimes be diagnosed years later, such as dyslexia that cannot be diagnosed before the age of formal reading instruction. For instance, Boets et al. (2011) showed an association between deficits in speech-in-noise perception at kindergarten, and diagnosis of dyslexia three years later, at 3rd grade. In their study, frequency modulation sensitivity, together with speech-in-noise

perception, uniquely predicted growth in reading ability. Guzek & Iwanicka-Pronicka (2022) showed the importance of screening for auditory processing disorders at 6 years of age and out of the three different auditory processing tests they used in a population of 1012 children (880 of whom presented diagnosed auditory processing disorder), speech perception in babble noise was the most frequently impaired. Here we investigate the development of speech perception in babble noise in children.

Indeed, one of the difficulties in implementing such a screening test in children, is the complex neurodevelopmental course of the many processes involved in speech perception in noise, with large inter-subject variability. Both the speech perception task and the type of noise masker have a major influence on the scores. Classically, energetic masking refers to noise maskers with the same spectro-temporal structure as the signal, hence reducing the availability of signal cues and affecting its neural representation, mostly at a peripheral level (e.g., white noise, speech-shaped noise). In contrast, informational masking refers typically to the situation of attending to one talker in the midst of another or several talkers (termed speech-on-speech masking, or babble noise or cocktail party noise), with masking occurring even when the target talker is perfectly audible. Many central factors are at play in those situations, such as the ability to selectively attend to the target talker whilst ignoring others, factors whose maturation go well beyond primary school years (Elliott, 1979). Most studies in children have used energetic masking (mostly speech-shaped noise) and pointed out a relative deficit, compared to adults, in speech perception in noise in 5 to 12 years old children, with speech in noise reception thresholds 3 to 7 dB greater than young adults (Fallon et al., 2000; Wilson et al., 2020). Such relative deficit, compared to adults, is all too often ignored and makes young children more susceptible to suffer detrimental effects of noise on speech

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

perception (Carvalho et al., 2017) and novel word learning (McMillan & Saffran, 2016). This deficit can seriously impair their formal learning in the classroom, whose acoustics are rarely adapted to their needs (Anderson, 2004; Finitzo & Tillman, 1978; Héту et al., 1990). Indeed, in active elementary classrooms, i.e., in a realistic environment, Sato & Bradley (2008) measured a mean signal-to-noise ratio (SNR) of 11 dB, ranging from 3 to 16 dB, and conducted speech tests. They established that, to reach 95% speech intelligibility, 6-year-olds needed a SNR of 15.5 dB whereas 11-year-olds needed 8.5 dB. Those SNR conditions for optimal speech perception were met only 20% of time for 6-year-olds and 49% for 11-year-olds (Bradley & Sato, 2008). Realistic classroom environments include the teacher's voice as the target signal, but all the teacher's voice reflections on the different surfaces (reverberation on tables, ceiling, walls...) end up as multiple speech signals originating from different positions in the auditory space, arriving at the child's ear with different delays. Short reverberation times (<0.5 s) have not been shown to be detrimental to word recognition, but longer delays induce temporal smearing of the acoustic information and partial masking, decreasing word recognition in the classroom environment (Wróblewski et al., 2012). Furthermore, reverberation times measured in classrooms are often above the recommended reverberation time of 0.6 s (Coffeen, 2000), sometimes up to 1.5s (Anderson, 2004; Crandell & Smaldino, 2000), rendering the type of noise closer to speech on speech. In addition to those multiple speech sources from the teacher, environmental noises and speech sources from other children within the class add up to form a noise masker closer to a multitalker babble noise than to energetic masking. Using a babble noise at SNR ranging from -4 to 4 dB, Howard et al. (2010) showed an increase in listening effort measured by a dual task paradigm on 31 school children aged 9 to 12 years old, with a decrease

in performance in the secondary task for increasing levels of noise, showing the importance of listening effort in a masker similar to actual classroom noise. In hearing-impaired children, speech in two-talker babble noise was closer to the receptive communication abilities of the children, reported by parents, than classical clinical tests in speech-shaped noise (Hillock-Dunn et al., 2015). Two-talker babble noise is known to be more challenging than speech-shaped noise, with, in young adults, more than 2 dB SNR difference in thresholds (Corbin et al., 2016; Lewis et al., 1988), and almost 6 dB in 5 to 7 year olds, in an open response test using monosyllabic words (Corbin et al., 2016). Further evidence suggested maturation of speech in multi-talker noise on a longer time frame than speech in speech-shaped noise (Corbin et al., 2016; Elliott, 1979; Hall et al., 2002). Indeed, speech in speech-shaped noise is supposed to be near adult values between 10 and 12 years of age (Koopmans et al., 2018). One of the first study of children's perception in babble noise showed increasing performance at repeating the final word of sentences, from 9-year-olds to 17-year-olds, with significant differences even between 15 and 17 year-olds, arguing for a long developmental time frame for speech in babble noise (Elliott, 1979). Calcus et al. (2018) showed a significant decrease in speech perception scores from a single-talker babble noise (94% at a fixed SNR of 0 dB) to a four-talker babble noise (64%), without significant difference between the four-talker and eight-talker babble noise (63%) in 14 French speaking children from elementary school. Interestingly, dyslexic children and children at risk for dyslexia showed consistently lower performance than control children in babble noise and in modulated speech noise, whereas no significant differences were obtained in quiet, nor in continuous speech noise.

Babble noise is considered to be the most challenging noise for speech perception and the closest to ecological conditions, and might be more sensitive

to listening difficulties associated with learning disorders than speech shaped noise. The number of talkers (Freyman et al., 2004; Rosen et al., 2013; Simpson & Cooke, 2005), the distance in fundamental frequency and the difference in language type (native versus non-native, Calandruccio et al., 2016; Calandruccio & Zhou, 2014) between the target talker and the multi-talker competitive babble, have all shown major influence on the masking efficiency in adults. As the number of competitive talkers increase from 1 to 6, consonant identification scores in young normally hearing adults decrease for up to 6 talkers, and then remain fairly stable or increase slightly from 8 to 128 talkers but remain under the scores obtained in a speech shaped noise (Simpson & Cooke, 2005). Rosen et al. (2013) obtained a sharp decrease in sentence identification scores from 1 to 2 talkers maskers, and a small increase of scores with 4 to 16 multi-talker babble, similar to Freyman et al. (2004)'s results, with a small increase of scores from 4 to 10 talkers in their multi-talker babble noise. Indeed, as the number of talkers increase, less gaps in the noise amplitude occur, hence less glimpsing opportunities that would help speech perception (Rosen et al., 2013). We would expect babble noise to resemble speech shaped noise beyond a certain number of talkers. However, even with 128 talkers, the babble noise is still more challenging than a speech shaped noise masker (Simpson & Cooke, 2005), showing that multi-talker babble has still a strong component of informational masking. Hence, we chose here to use a 16-talker babble noise, in order to avoid the greater variability linked to few-talkers maskers and being closer to realistic listening situations, involving both energetic and informational masking.

However, defining the type of masker to use is only part of the story to design a speech-in-noise test adapted for children. Indeed, most previous studies use open-set response paradigms, which can limit their application in children.

For instance, in the first “babble noise in children study”, Elliott (1979) used sentences with different predictability of words, and the lower age limit considered for such test was 9 years of age. Open-set speech tests, that require verbal responses, need clear articulation by the child to allow precise scoring (Stiles et al. 2012), which can make them problematic for hearing-impaired young children, children with disorders of speech sound production (Cabbage & Hitchcock, 2022), or children whose verbal responses can be difficult to obtain (for instance in some children with non-speaking autism). By using picture-pointing response requiring to choose an answer amongst several foils, alternative forced choice tests (nAFC) alleviate the need for a verbal response and are very valuable in paediatric assessments (Mendel, 2008; Vickers et al., 2018). Furthermore, word recognition is heavily dependent on psycholinguistic factors, with more familiar words and words with high occurrence frequency being easier to recognize (Howes, 1957; Savin, 1963). Occurrence frequency influences word recognition in conjunction with phonological neighbourhood: indeed, words that differ, in the lexicon, from a lot of other words by only one phoneme (phonological neighbours), are more difficult to identify than words which do not have a lot of phonological neighbours. Furthermore, if those phonological neighbours are not only numerous, but of greater occurrence frequency than the target word, this target word is even more difficult to recognize (Luce & Pisoni, 1998). This allows to categorize words into “easy words” and “hard words”, and the influence of such psycholinguistic factor on open-set word recognition has been largely shown (Dirks et al., 2001; Kirk et al., 1997; Meyer & Pisoni, 1999), including in children (Kirk et al., 1995; Krull et al., 2010). One way to decrease the importance of such psycholinguistic factors is to use closed-set tests. Indeed, as the subject is presented with the potential answers and therefore doesn’t have to access a potentially unlimited

amount of words in their mental lexicon during the task, the closed-set format doesn't reflect the many complex processing features involved in word recognition in naturalistic environments, but rather reflects auditory abilities per se. By being less dependent on the internal lexicon development, the closed set format could help eliminating a substantial portion of unwanted variability in the processing of the signal, especially in children, who show a large variability in the extend and organization of their mental lexicon with age. Indeed, the independence of test scores from vocabulary knowledge is one of the recommended criteria for behavioral tests of speech perception in children (Kosky & Boothroyd, 2003; Mendel, 2008) and tests should assess perceptual capacity, rather than more general cognitive abilities (Hnath-Chisolm et al., 1998). Comparing a 6AFC closed format test with an open response one, Sommers et al. (1997) showed, in adults, the absence of lexical difficulty influence on 6AFC scores, whether the 6AFC task involved words that are phonetically similar to many other words (hence a more demanding task) or words with few phonetically similar competitors.

In the present study, we used a French language adaptation for children of the English versions of 4AFC tests (Buss et al., 2016; Vickers et al., 2018), using pictures easily recognizable by 5-year-olds, and two different phonological conditions. A "difficult" condition, where words were phonologically similar to one another and therefore easily confusable, and an easy condition, where words were very dissimilar. In order to make the two conditions strictly comparable, the same targets were presented in both conditions, in a randomized manner, using fixed SNRs in separate blocks. In addition, the use of this fixed SNR paradigm allowed us to assess the potential existence of a training effect with repetition of the task, and the potential influence of psycholinguistic factors. Indeed, this would have been difficult to assess using an adaptative method

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

that involves trial per trial modifications of the SNR and has been described as more sensitive to fatigue and inattention (Wightman & Allen, 1992). Lastly, the use of a touch tablet allowed us to take advantage of the greater motivational aspect of such computerized tests in children, and gave us easy access to response time, in addition to accuracy data. This allowed us to investigate, in a systematic manner, developmental aspects of speech perception in babble noise, in 5 to 10 years old children, relative to a group of young normal-hearing adults.

Methods

Participants

One hundred children (mean age = 7.6 years; min = 4.5 years, max = 10.6 years, 6 left-handed) attending a public primary school in southeast France participated in this study, as well as in a companion study on the development of auditory short-term memory (Ginzburg et al., 2022). The children were tested during school hours. Participation in the study was proposed to a total of 114 children and they were included in the study only if both parents or legal tutors provided a written informed consent. The study was approved by the relevant services of the French public education services (Inspection de l'Education Nationale (IEN) and Direction Académique des Services de l'Education Nationale (DASEN) of the Isere department). Children from Kindergarten (KG) grade to 5th grade performed the experiment. Before testing, parents filled a questionnaire about their child's level of education and their own level of education (data reported in Ginzburg et al., 2022), the child's laterality, possible vision or auditory impairments, musical activities, bilingualism, learning disabilities, and 11 questions adapted from adults' musical listening questionnaires (Lévêque et al., 2018; Tillmann et al., 2014). Among the 100 children, parents' responses to questionnaires revealed that 5

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

children had a diagnosed learning disability and their data were excluded from the current analysis. Thirty children had seen at least once a speech-therapist and 6 had already worn ear tubes. As we aimed here to explore the cognitive abilities of children among a representative set of the population, we present the results including data from all these children. Data from five children were excluded because of technical difficulties during recordings. Overall, data from 10 children were thus excluded from the analysis, the sample size of the remaining 90 children per grade are reported in Table 1. Data on participant ethnicity were not collected because the collection of such data is illegal in France and would require an exceptional waiver from government agencies that we did not seek.

Nineteen adults (four men), with a mean age of 24.11 years ($SD = 3$ years) were also included in the study. None of them presented any visual or auditory impairment, nor neurological or psychiatric troubles, and none of them had any musical background.

English label	Kindergarten (KG)	1 st Grade	2 nd Grade	3 rd Grade	4 th Grade	5 th Grade
French label	Grande Section de Maternelle (GS)	Cours Préparatoire (CP)	Cours Élémentaire 1 (CE1)	Cours Élémentaire 2 (CE2)	Cours Moyen 1 (CM1)	Cours Moyen 2 (CM2)
N	11	16	17	16	15	15
Mean age in years (SD)	5.01 (0.31)	6.03 (0.29)	6.93 (0.30)	8.01 (0.34)	8.92 (0.26)	9.94 (0.29)

Table 1 : Number of participants and mean age (SD in parentheses) for each grade and matching English and French labels for educational level.

Stimulus construction and task design

Each trial of the speech-in-babble task consisted in matching an aurally-presented word with its corresponding image among four other images. The material consisted of 20 monosyllabic and 4 disyllabic words, easily representable by a drawing, selected as a function of their concreteness, their frequency of occurrence in the French language, their age of acquisition, and their phonological similarity. The 24 selected words were recorded by a French native female talker, in a sound proof booth, using a Rode NT1 microphone. Sound files had an average duration of 540 ms (SD=171 ms) and their amplitudes were equated in root mean square (RMS) level. Each word had a corresponding hand-drawn image. From this list of 24 words, six sets of four phonologically close words were created for the difficult condition, e.g., /ku/, /kul/, /ru/, /ruʒ/ (cou, coule, roue, rouge in French, corresponding to neck, flow, wheel, red in English). Within each difficult set, only consonants could change and up to three phonemes were modified. With the same 24 words, six sets of four phonologically distant words were created for the easy condition, e.g., /bɑ̃/, /flœʁ/, /mɛ̃/, /trwa/ (banc, fleur, main, trois in French, corresponding to bench, flower, hand, three in English), where several phonemes, including the vowels, were different. The average of pairwise Levenshtein distance (i.e., the minimal possible number of insertions, deletions, and substitutions needed to transform one chain of phonemes into another) was 3.03 phonemes (SD = 0.4) for the easy condition and 1.39 phonemes (SD = 0.29) for the difficult condition. Using the French database ‘Lexique’ (New et al., 2004), we calculated the number of phonological neighbors (i.e. number of existing words by changing one phoneme without changing the others) for each of the 24 words used, and then extracted the number of unique phonological neighbors for each 4-word set. In the easy condition, the four words of each set

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

were phonologically distant (each word having phonological neighbors different from the other three) thus the number of unique phonological neighbors within a set was relatively high (mean = 83, sd = 6.7). Conversely, in the difficult condition, the four words were quasi-phonological neighbors (each word in a set having several common phonological neighbors with the other three) thus having a relatively small number of unique phonological neighbors within a set (mean = 59.7, sd = 14.8). Lastly, we derived for each of the target words their frequency of occurrence (per million occurrences) in the French language (mean = 85.9, SD=119) calculated from a movie subtitles corpus (New et al., 2007) (for details about the stimuli, see supplemental Table S1 and Table S2).

Procedure

Children were tested by groups of five or six in the sport room of their school, after performing auditory short-term memory tasks (data reported in Ginzburg et al., 2022). Before testing, children sat in front of their tables, listening to the experimenter's instructions. The experimenter explained the speech-in-babble noise task with cardboard signs: children would see 4 images appear on the touchpad and at the same time, they would hear a word through their headphones. They had to find the spoken word in one of the four images and tap on it as quickly as possible. They had to ignore the people talking in the background (babble noise). The task was implemented on a touch-tablet and auditory stimuli were displayed binaurally through Sennheiser HD-250-Pro circum-aural headphones. Children underwent a training block comprising eight trials with a SNR of +3 dB. Then they underwent two 48-trial blocks with a SNR of +3 dB for the first block and a SNR of -3 dB for the second block. The target-word was systematically presented at 66 dB SPL (A-weighted). During each trial, one of the four-image set was presented, and 800 ms later, the target word was aurally presented. The subject had to tap with a finger on the

matching image as quickly as possible but without any time limit. The next trial began immediately after the response. In both blocks, twelve four-image sets were used: six consisting of phonologically distant words (easy condition) and six consisting of phonologically close words (difficult condition, see above). Within a block, each set was presented four times, each time with a different target word (since the same words were used to create the easy and the difficult condition, each word was used twice as the target: once in each of the two phonological difficulty condition). A block was thus divided in four series of 12 trials with each of the four-image set presented once per series. The order of presentation of the 12 sets was randomized for each subject and each block, and the same randomized order was then used for the four series. This was done to avoid, as much as possible, the potential advantage linked to the limitation of the number of remaining response alternatives once items have been eliminated from the list of possibilities during the first series. The position of the images presented on the screen was randomized. Lastly, no feedback was given on a trial per trial basis. During the task, a babble noise made of 16 unintelligible French native male and female voices (Moulin et al., 2013) was presented in a continuous manner. The subjects could listen to several seconds of the babble noise before themselves, triggering the start of the block.

The duration of each block was around 2 minutes, so that the total hearing-in-noise test was performed in about 10 minutes briefing and training included. The worst SNR condition was always given last, as it was anticipated that the children listening behavior might be modified by the challenge represented by this difficult SNR (i.e., inducing discouragement before the next block).

Adults underwent the task individually in an experimental room in the lab with the same equipment as children. They performed the same eight-trial training block as the children and then eight 48-trial blocks with babble noise ranging

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

from -12 dB SNR to +9 dB SNR by steps of 3 dB SNR. The order of the blocks was counterbalanced.

Data analyses

All data analysis were performed on R (R Core Team, 2019, sample code is available at https://github.com/jeremieginzburg/cd_analysis_sample_code). For each participant, accuracy (percentage of correct responses) and response times (RT) from sound onset were extracted for each phonological condition and for each SNR. Trials with RTs above 8 seconds and when the participant answered before the onset of the sound were excluded, representing 1.4% of children's data and 0.1% of adult's data. For all analysis on RTs, we used response times on correct trials only and on trials with common words between the easy and the difficulty condition (to avoid imbalance due to the fact that more trials were successful in easy than in difficult condition), which represented 75.4% of the data for children and 93.5% for adults.

Adult data: psychometric parameters

To evaluate if detection thresholds differed according to the phonological difficulty of the task, psychometric curves were fitted to responses as a function of the signal-to-noise ratio using a logistic function with a guess rate λ set to .25 (because of the 4AFC task's nature) for each difficulty condition and for each participant by using the *quickpsy* R-package (Linares & López-Moliner, 2016). It allowed the estimation of detection thresholds (i.e., SNR for which each participant obtains $\lambda + (1 - \lambda) * 0.5 = 62.5\%$ of correct responses thus half-way between chance level (25%) and perfect performance (100%)). Easy and difficult conditions were compared for each subject with the *thresholdcomparisons* function from the *quickpsy* R-package which performs a bootstrap comparison between conditions to determine if the difference between

them falls outside a 95% confidence interval. Comparison between conditions at the group level was tested by comparing the average detection thresholds across all subjects for both difficulty conditions with a pairwise t-test after checking for normality with a Shapiro-Wilk test.

Development of hearing in babble noise

Developmental aspects of children's performance were tested as a function of children's school grade, allowing for homogeneous groups composed of children being equally scholarly educated regarding reading and other abilities. In order to assess the effects of grade, phonological difficulty, and SNR on children's performance, we used general linear models (GLM) for accuracy and general linear mixed models (GLMM) for response times (Bates et al., 2014). We thus analyzed the influence of three main effects on percent correct scores and RTs:

- Between-subject factor *Grade*: six levels (KG, 1st, 2nd, 3rd, 4th and 5th grade)
- Within-subject factor *Difficulty*: two levels (easy/difficult phonological condition)
- Within-subject factor *SNR*: two levels (-3 dB/+3 dB)

GLM(M) were fitted using the *lme4* R package (Bates et al., 2014, p. 4). Statistical tests were performed using type II Wald chi-square analysis of variance using the *car* package (Fox & Weisberg, 2018), all effect-sizes were calculated for each fixed effect with the *esc* package (Lüdtke, 2019).

When a significant main effect or interaction was found, estimated marginal means post-hoc tests were performed using the *emmeans* package (Lenth, 2021) and corrected for multiple comparisons using the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995).

Two measures were considered to investigate children's performance:

- *Accuracy*: percent correct scores were fitted with a binomial distribution as GLMs can handle non-normally distributed data and, in particular, count data (Bates et al., 2014). Data were thus fitted with a binomial distribution with a logit link function using the *glm* function from the core *stats* package. We did not use mixed models for the accuracy because one proportion of correct response was obtained for each condition and each subject, thus rendering the intra-subject variability null.
- *Response times*: As recommended by Lo and Andrews (2015), RTs were not log-transformed to fit a gaussian distribution which can produce additive effects. Since raw RTs distribution can be fitted either with an inverse gaussian or a gamma distribution, we compared four models, fitted with either a gamma or an inverse gaussian distribution and with either an identity link function or a log-link function. The model with the lowest AIC was then chosen. The random-effects structure was built according to Bates et al. (2015)'s specifications: we used the subject variable as an intercept and we started by including all within-subject fixed-effects as random slopes (i.e., all effects were allowed to vary across subjects) and dropped slopes factors one by one. After rejecting the non-converging models, we chose the model with the lowest AIC.

We also compared measures between the oldest children (5th grade) and adults. We extracted adults' performance for the two SNR conditions that children underwent (-3 and +3 dB SNR) and we performed similar GLM(M)s than the previously described ones (on accuracy and RTs), replacing the *Grade* factor by a *Group* factor with two levels: 5th grade and adults. We kept the *difficulty* and *SNR* factors.

Impact of parameters potentially influencing performance

- Repetition and learning: Influence of trials repetition within a block

The effect of trials repetition within a block was assessed by analyzing children's accuracy (percentage of correct responses) as a function of the series of 4-image sets presented during a block (as each block consisted of four consecutive 12-trial series). Accuracies were calculated per subject, per block, and per series and were fitted using the *glm* function from the core *stats* package with a binomial distribution and a logit link function with grade (KG to 5th grade), difficulty (easy/difficult), SNR (-3 dB and +3 dB) and series (1 to 4) as factors. In addition, the reproducibility within a block was assessed by a two-way random effect interclass correlation.

- Word/picture familiarity

To check the potential changes in word/picture familiarity across the different grades, we ranked the 24 target words according to average accuracy (percentage of correct responses) for each age group (highest score: rank=1, lowest score: rank=24) for both easy and difficult conditions, using the same method as Fallon et al. (2000). We then performed ranked correlations (Spearman method) between the ranks for the different age groups, and between the rank of each age group and the overall average ranking.

- Psycholinguistic parameters

The effect of two psycholinguistic factors (word frequency of occurrence and number of phonological neighbors, see above) calculated from the Lexique database (New et al., 2004), was assessed on children's accuracy. For frequency of occurrence, mean accuracy was calculated for each target-word (24 in total), each difficulty (easy/difficult) and each SNR (-3/+3 dB SNR). A GLMM was fitted on proportion data (number of correct responses over number of trials)

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

and fitted with a binomial distribution, a logit link function and target-word as random factor. In each GLMM, the effect of the frequency of occurrence factor was tested in addition with the grade factor (six levels: kindergarten to 5th grade), the difficulty factor (easy/difficult) and the SNR factor (-3/+3 dB SNR).

For phonological neighbors, mean accuracy was calculated for each four-image set (six sets of the easy condition, six sets for the difficult condition). One GLMM was fitted for each difficulty level, as by construction, the number of unique phonological neighbors for a set is smaller for easy sets than difficult sets. These two GLMMs were performed on accuracy and fitted with a binomial distribution, a logit link function, and four-image set as random factor. In each GLMM, the effect of the phonological neighbors factor was tested in addition with the grade factor (six levels: kindergarten to 5th grade) and the SNR factor (-3/+3 dB SNR). Statistical tests were performed using type II Wald chi-square analysis of variance using the *car* package (Fox & Weisberg, 2018).

Results

Adults data: psychometric parameters

Detection thresholds were computed for each participant and for each phonological condition. At the individual level, 74% of the adult participants displayed significantly different detection thresholds between the easy and the difficult conditions (fourteen out of the nineteen participants, Figure 1a). At the group level, detection thresholds averaged across participants were significantly different between the easy and the difficult condition (mean threshold for the easy condition = -10.6 dB; mean threshold for the difficult condition = -6.3 dB; $t(18) = 10.57$, $p < .001$, Figure 1b). Thus, in adults, to achieve similar performance in easy and difficult phonological conditions, an

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

increase of ~ 4.3 dB of SNR is necessary in the difficult condition relative to the easy one.

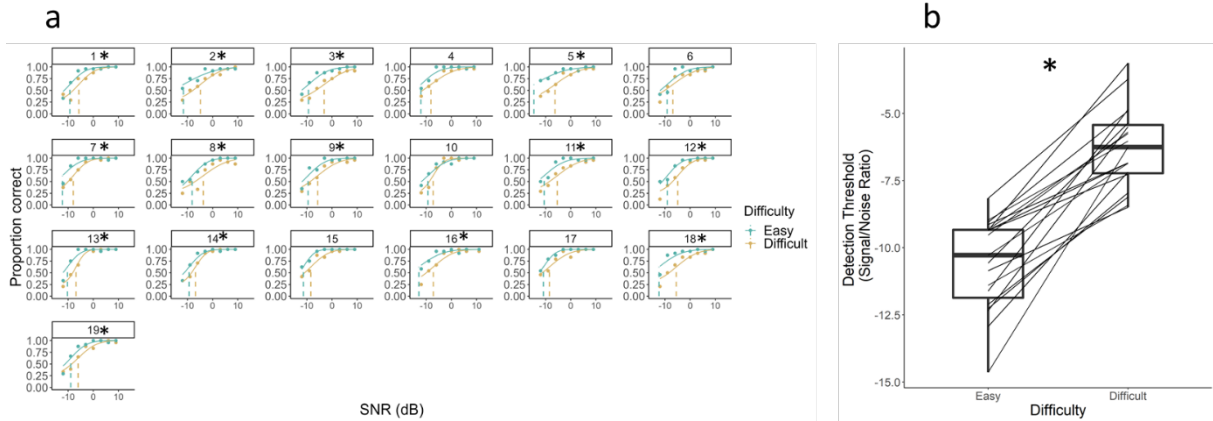


Figure 1: subject-level and group-level effect of the phonological condition on psychometric thresholds in adults. a: fitted psychometric functions on the proportion of correct response for each subject (numbered panels) and each phonological condition (blue lines: easy/orange lines: difficult) as a function of the SNR during the task. Dotted lines represent the projection of the estimated SNR threshold needed to achieve 62.5% correct response for each condition. Asterisks above the participant number indicate significantly different thresholds for the two conditions at the individual level. b: average detection thresholds across participants as a function of the phonological condition. Individual data are represented by black lines. Asterisk indicate a significant difference between the two conditions.

Development of hearing-in-babble noise

GLM(M) were performed on accuracy data and response times with Grade (kindergarten to 5th grade), Difficulty (easy/difficult), and SNR (-3 dB/+3 dB) as factors. Mean accuracy and response times are shown in Figure 2. Complete statistical results are shown in Table 2 and we present significant effects, interactions, and their follow-up post-hoc tests below.

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

<i>Measure</i>	<i>Factor</i>	<i>Df</i>	χ^2	<i>p</i>	<i>Cohen's d</i>
Accuracy	Grade	5	22.9	3.5 - 04	1.168
	Difficulty	1	61.9	3.62e - 15	2.968
	SNR	1	32.02	1.53e - 08	1.486
	Grade:Difficulty	5	0.5	0.992	0.149
	Grade:SNR	5	1.3	0.935	0.242
	Difficulty:SNR	1	2.43	0.119	0.333
	Grade:Difficulty:SNR	5	0.83	0.975	0.193
Response times	Grade	5	12.901	0.0243	0.489
	Difficulty	1	9.491	0.00206	0.687
	SNR	1	0.008	0.928	0.02
	Grade:Difficulty	5	3.428	0.63	0.10
	Grade:SNR	5	3.523	0.62	0.10
	Difficulty:SNR	1	0.404	0.525	0.134
	Grade:Difficulty:SNR	5	0.39	0.99	0.0012

Table 2 : Type II Wald chi-square analysis of deviance results on the GLM with binomial modeling for the accuracy measure and on the GLMM with inverse gaussian modeling for the correct response time measure for children data. All main effects were tested for each measure: Grade (six levels: KG to 5th grade), Difficulty (two levels: easy/difficult), SNR (two levels: -3 dB/+3 dB) and their interactions (see text for details). Significant effects are depicted in bold font (p -values < .05). df : degrees of freedom. χ^2 : chi-square test. p : p -value. Cohen's d : standardized mean difference.

Accuracy. The following model was used for accuracy: $correct\ responses/number\ of\ trials \sim grade^* SNR^* difficulty^*$, family =

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

binomial(link=logit). The main effect of Grade was significant ($\chi^2(5) = 22.9$, $p < .001$, $d = 1.17$) revealing better performance for older children. Post-hoc tests revealed a significantly lower performance for KG children compared to 3rd, 4th, and 5th graders ($p < .0029$), significantly lower performance for 1st graders compared to 5th graders ($p < .03$), marginally significant lower performance for 2nd graders compared to 5th graders ($p = .06$). The main effect of Difficulty was significant ($\chi^2(1) = 61.9$, $p < .001$, $d = 2.97$) with lower performance in the difficult condition. The main effect of SNR was also significant ($\chi^2(1) = 32.02$, $p < .001$, $d = 1.49$) with lower performance for the lowest SNR (-3 dB). No interaction effect was significant ($p > .12$).

Response times. The model that best fitted our data was the one using an inverse gaussian distribution with a log-link function, with random effect structure including Difficulty and SNR as slopes (*response times* ~ *grade*difficulty*SNR* (*SNR* + *difficulty* / *subject*), *family=inverse.gaussian(link=log)*). The main effect of grade was significant ($\chi^2(5) = 12.901$, $p = .024$, $d = .49$). Post-hoc tests only revealed a marginally significantly faster response time for 5th graders compared to KG ($p = .05$) and for 5th graders compared to 2nd graders ($p = .0503$). Interestingly, longer response times were obtained for the easy condition compared to the difficult condition (main effect of Difficulty ($\chi^2(1) = 9.491$, $p = .002$, $d = .69$)). No effect of SNR or interaction effects were found ($p > .53$).

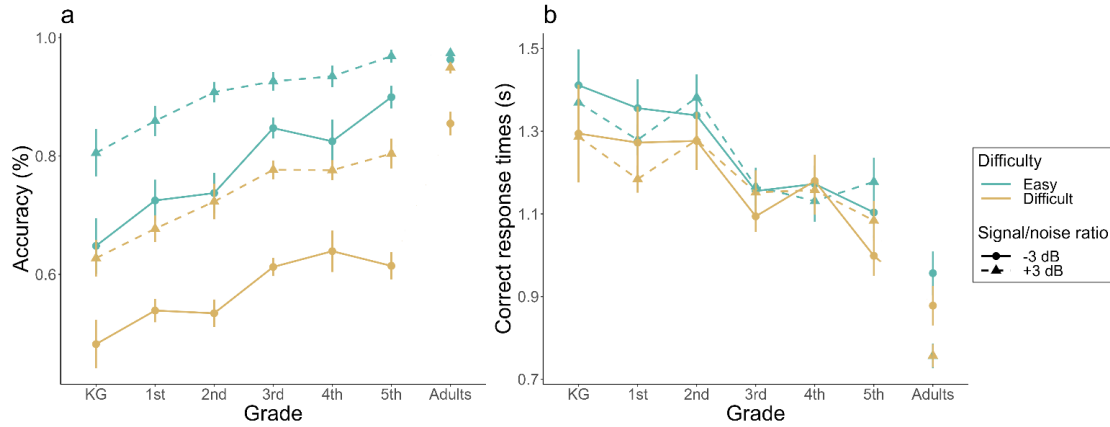


Figure 2: (a) mean performance (% of correct response) and (b) mean correct response times after onset of the sound (in seconds) as a function of SNR and Difficulty conditions for all grades and for adults. Response times are for words correctly recognized in both easy and difficult conditions (per participant). Error bars represent standard deviations around the mean. Blue color represents the easy condition; orange color represents the difficult condition. Solid lines represent the 3dB SNR condition, dashed lines the -3 dB SNR condition. KG: Kindergarten. Adult's average response times for the easy and difficult condition in 3 dB SNR were similar, so that in the figure (b), the triangle for the easy condition hides the one for the difficult condition.

Comparison between children and adults' performance

Comparison of children's performance and adult's psychometric parameters

In order to obtain a representative adult psychometric curve, we fitted a psychometric curve to accuracy averaged across adult data separately for the easy and the difficult conditions (respectively left and right panels in Figure 3). We then projected children's performance on the overall adult psychometric curve as a function of their grade and of the SNR used (-3 dB and +3 dB, respectively top and bottom panels in Figure 3). We thus obtained the value of the SNR at which an adult would have to be tested to obtain the same performance as children. This representation gave us an idea of the maturation gap between children and adults for their speech-in-babble-noise abilities. Indeed, the « SNR gap » (i.e., the difference between the SNR used to test children and the SNR corresponding to the same performance in adult data)

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

goes up to 10 dB for the youngest children and the minimal one is 1 dB for the oldest children in the easy condition and for a testing SNR of -3 dB.

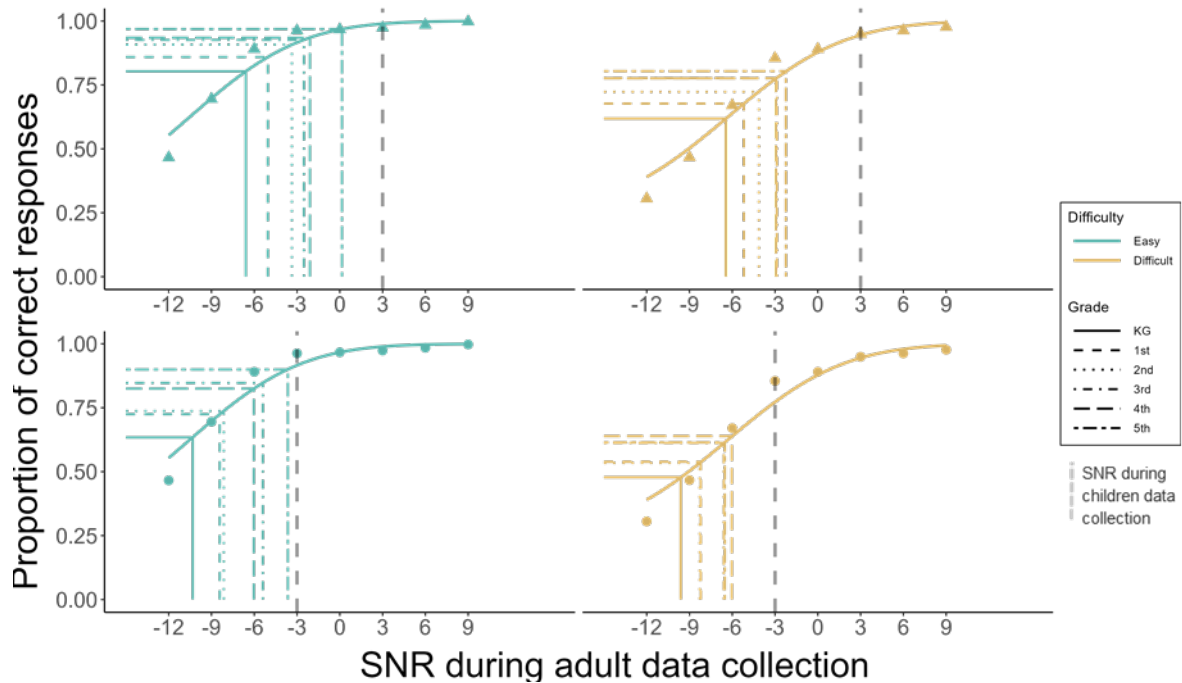


Figure 3: projection of children's performance by grade on the adult's psychometric function (derived from the average data of the 19 adult participants). Left panels and blue lines for the easy condition; right panels and orange lines for the difficult condition. Top panels: the projected performance of children when they underwent the -3 dB SNR condition. Bottom panels: the projected performance of children when they underwent the +3 dB SNR condition.

Comparison between performance of oldest children and performance of adults

To compare performance of the oldest children (5th grade) with performance of adults, we performed GLM(M)s on the same measures as for children (accuracy and response times) with Group (adults (n = 19) and children (5th graders, n = 15)), Difficulty (easy/difficult), and SNR (-3dB/+3dB) as factors. Mean accuracy and response times are shown in Figure 2. Complete statistical results are shown in Table 3 and we present significant effects, interactions, and their follow-up post-hoc tests below.

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

<i>Measure</i>	<i>Factor</i>	<i>Df</i>	χ^2	<i>p</i>	<i>Cohen's d</i>
Accuracy	Group	1	6.159	.0131	0.941
	Difficulty	1	9.547	.002	1.25
	SNR	1	4.044	.0443	0.735
	Group:Difficulty	1	4.079	.0434	0.738
	Group:SNR	1	1.19	0.275	0.38
	Difficulty:SNR	1	1.931	.165	0.491
	Group:Difficulty:SNR	1	.319	.572	0.195
Response times	Group	1	26.919	2.12e-07	3.9
	Difficulty	1	3.199	.074	0.531
	SNR	1	2.241	.134	0.645
	Group:Difficulty	1	2.309	.129	0.54
	Group:SNR	1	12.625	3.8 -04	1.537
	Difficulty:SNR	1	1.125	.289	0.37
	Group:Difficulty:SNR	1	0.002	.965	0.0152

Table 3 : Type II Wald chi-square analysis of deviance results on the GLM with binomial modeling for the accuracy measure and on the GLMM with inverse gaussian modeling for the correct response times measure, to compare data in the oldest children with adult data. All main effects were tested for each measure: Group (two levels: 5th grade vs. adults), Difficulty (two levels: easy/difficult), SNR (two levels: -3 dB/+3 dB) and their interactions (see text for details). Significant effects are depicted in bold font (*p*-values < .05). *df*: degrees of freedom. χ^2 : chi-square test. *p*: *p*-value. Cohen's *d*: standardized mean difference.

Accuracy. The following model was used for accuracy: *correct responses/number of trials* ~ *group*difficulty*SNR*, family = *binomial(link=logit)*). All main effects were significant: Group ($\chi^2(1) = 6.159$,

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

$p = .0131$, $d = 0.94$ with adults having better performance than 5th graders, Difficulty ($\chi^2(1) = 9.547$, $p = .002$; $d = 1.25$) with better performance in the easy condition, and SNR ($\chi^2(1) = 4.044$, $p = .044$, $d = .74$) with better performance for the +3 dB SNR. In addition, a significant Group-by-Difficulty interaction was found ($\chi^2(1) = 4.079$, $p = .043$, $d = .74$) with significantly better performance in the easy condition compared to the difficult one for 5th graders ($p < .001$) but not for adults ($p = .29$).

Response times. The model that best fitted our data was the one using an inverse gaussian distribution with a log-link function, and a random effect structure including Difficulty and SNR as slopes (*response times* ~ *group*difficulty*SNR + (SNR + difficulty / subject), family = inverse.gaussian(link=log)*). There was a marginally significant effect of Difficulty ($\chi^2(1) = 3.199$, $p = .074$, $d = .65$) for which longer response times were found for the easy condition compared to the difficult condition. Critically, the main effect of Group was significant ($\chi^2(1) = 26.919$, $p < .001$, $d = 3.9$) as well as the Group-by-SNR interaction ($\chi^2(1) = 12.625$, $p < .001$, $d = 1.54$), that revealed longer response times for the lowest SNR (-3 dB) for children ($p = .018$) but not for adults ($p = .66$).

Impact of learning, familiarity, and psycholinguistic parameters

Learning in the course of a block: Influence of the repetition of the trials

The model used in this analysis of accuracy is the following: *correct responses/number of trials* ~ *grade*difficulty*SNR*series, family = binomial(link=logit)*). As expected from the developmental trajectory analysis above, effects of Grade ($\chi^2(5) = 22.90$, $p < .001$, $d = 1.17$), Difficulty ($\chi^2(5) = 62.19$, $p < .001$, $d = 2.99$) and SNR ($\chi^2(5) = 32.18$, $p < .001$, $d = 1.49$) were found. No effect of series was observed ($\chi^2(3) = .32$, $p = .96$, $d = .12$) indicating

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

an absence of significant increase or decrease of performance within a block. No interaction effect was found (all $p > .45$). Reproducibility was good with an interclass correlation coefficient of .60 ($F = 2.57$, $p < .001$) for the SNR -3 dB condition and of .68 ($F = 3.14$, $p < .001$) for the SNR +3 dB condition.

Word picture familiarity

Ranking of performance per word were correlated between groups, and all pairwise correlations were significant ($p < .001$, r ranging from .693 to .867). Correlations between the ranking of each group with the overall ranking were also significant ($p < .001$, r ranging from .827 (KG) to .916 (2nd grade).

Influence of psycholinguistic parameters

Frequency of occurrence

The effect of the Frequency of occurrence of target words, Grade (kindergarten to 5th grade), Difficulty (easy/difficult), and SNR (-3/+3 dB SNR) factors was tested on accuracy. Except for the expected Grade ($\chi^2(5) = 24.301$, $p < .001$, $d = .86$), Difficulty ($\chi^2(1) = 66.738$, $p < .001$, $d = 3.39$), and SNR ($\chi^2(1) = 34.608$, $p < .001$, $d = 1.58$) effects (already found in the previous analysis), no other significant effect or interaction was found (all $p > .11$), with, in particular, no significant effect of the frequency of occurrence.

Phonological neighbors

The effect of the Number of unique phonological neighbors within each 4-images set, Grade (kindergarten to 5th grade), and SNR (-3/+3 dB SNR) factors were tested on accuracy. One model was fitted for each difficulty condition (easy/difficult). The expected effect of Grade was found for the easy ($\chi^2(5) = 11.158$, $p = .048$, $d = .43$) and the difficult conditions ($\chi^2(5) = 12.485$, $p = .029$, $d = .47$), as well as the SNR effect for the easy ($\chi^2(1) = 9.548$, $p = .002$,

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

$d = .69$) and the difficult ($\chi^2(1) = 26, p < .001, d = 1.27$) conditions. All other effects or interactions were non-significant ($p > .11$).

Discussion

The present study depicts the neurodevelopmental aspects of speech perception in multi-talker babble noise, in French language, in a large population of primary school children, using a newly developed French closed-set test with a 4AFC paradigm. Additionally, the exploration of children's response times allowed to unveil an effect of competition between phonologically similar words within the lexicon. Finally, we explored the impact on performance of several factors that have implications for the design of central hearing screening test: reproducibility, word familiarity, and psycholinguistic factors.

Developmental trajectory of speech-in-babble-noise perception

The main results show definite steps into the neurodevelopment, with significant differences in accuracy between the youngest children (5-year-olds) and the older children (7-, 9-, and 10-year-olds), between 6-year-olds and 9-10-year-olds, and between 7- and 10-year-olds, revealing a major maturation step around 7 years of age. The accuracy obtained by the oldest children (10-year-olds) were equivalent to a difference of more than 3 dB SNR on average with normally hearing young adults, showing that speech perception in one of the most common, but challenging situation (i.e., speech-in-babble noise), is far from mature during nearly the entirety of primary and secondary school attendance years. Those results are in agreement with Wilson et al. (2010)'s large normative study, using an open set task in babble noise (Word in Noise test). Indeed, the authors showed that the largest improvement in recognition performance occurred between 6- and 7-year-olds, then a slight improvement between 7 and 9-year-olds, and performance remained fairly constant between

9- and 12-year-olds, with a significant difference between 12-year-olds and young adults.

It could be argued that phonological awareness maturation could account for the lower scores of the youngest children, especially for Kindergarten children. For instance, using an oddity non-sense syllable recognition task (presented with 3 auditory non-sense syllables - two identical and one odd), Hnath-Chisolm et al. (1998) obtained significantly lower accuracy for 5- to 7-year-olds compared to 7- to-9-, or 9-to-11-year-olds. Those differences could possibly be attributed to cognitive and phonological development for children under 7 years (Hnath-Chisolm et al., 1998). In the present study, even though we obtained the expected greater accuracy for phonologically distant words than for close words, we did not observe any interaction between grade and difficulty in the children's data analysis. This fairly parallel evolution of accuracy in easy versus difficult phonological conditions with age (see also Figure 4) argues against an important influence of phonological awareness maturation in the evolution of accuracy as a function of age, at least between 5 and 10 years of age. However, we observed a weakly significant interaction between difficulty and group when comparing children's data with adults' data (with a difference between difficult and easy conditions in children only). We thus cannot rule out the possibility of the effect of later phonological awareness maturation on the current test, although the absence of difficulty effect in adults could also be explained by a ceiling effect, driven by the easiest SNR condition (+3 dB SNR).

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

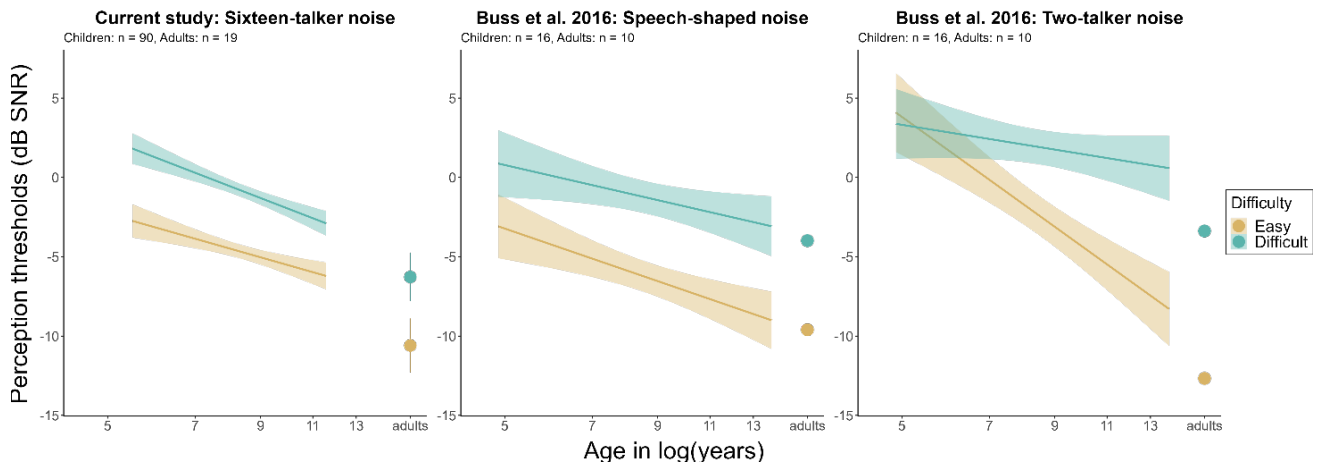


Figure 4 : comparison between the current study and Buss et al. (2016). For the current study a SNR threshold was derived for each child and each condition from the 2 points obtained along the psychometric curve and adding 2 other points: one at 100% correct score for a 15 dB SNR and one at 25% chance level for a -15 dB SNR. We then fitted a linear regression as a function of the decimal logarithm of age in years, as in Buss et al. (2022). For the Buss et al. (2016)'s figures, data were recovered from their Figure 2 using the Engauge Digitizer software (<https://markumitchell.github.io/engauge-digitizer/>). For the current study, the effect of age was significant ($p < .001$) and slopes were $-.53$ and $-.75$ for easy and difficult conditions respectively. For Buss et al.'s speech-shaped noise, the effect of age was significant ($p < .014$) and slopes were $-.64$ and $-.47$ for the easy and difficulty conditions respectively. For the two-talker noise, the effect of age was significant only for the easy condition ($p < .001$) but not for the difficult condition ($p = .14$). Slopes were -1.43 and $-.03$.

The difference between phonologically similar and dissimilar words obtained here (as measured with the procedure described in Figure 4, e.g., a threshold difference of 3.8 dB SNR in children and 4.3 dB in adults) appears to be lower than the one obtained by Buss et al. (2016) in a 4AFC task, in a speech shaped noise (about 5 dB in children and 7 dB in adults). Differences in language (French versus English), in the degree of similarity between the target words and the foils, and in the type of noise used, could account for that. In addition, Buss et al. (2016) used the 2 conditions in 2 different blocks, determining the SNR threshold using an adaptative method, whereas we used a fixed SNR method, where easy target words and difficult trials were randomly intermixed. Both in the present study and in Buss et al.'s data (Figure 4), performance improved with age for both easy and difficult phonological conditions. It is

interesting to note though, that our results match better with Buss et al. (2016)'s results in speech shaped noise than in their results in 2-talker masker. Indeed, in the 2-talker masker, they obtained a significantly sharper improvement of accuracy with age in the easy condition than in the difficult condition, with a difference of almost 10 dB SNR between both conditions in adults. This similarity between Buss et al. (2016)'s results in speech shaped noise and ours can be easily attributed to the 16-talker babble noise we used, that involves both energetic and informational masking, and is closer, in performance, to speech shaped noise than the strong informational masking of a 2 talker noise (Rosen et al., 2013). The 16-talker babble noise offers a good mix between energetic and informational masking. The definite existence of an informational masking component, involving central auditory processing, in this 16-talker babble noise, has been shown in a case of venous cerebral infarct in an adult patient (Bourgeois-Vionnet et al., 2020). Indeed, this patient showed a specific deficit in the 16-talker babble noise speech perception in the ear contralateral to the infarct, without any deficit in a speech shaped noise of the same spectrum, using the same test as in the present study.

The informational masking component could explain the absence of mature performance at 10 years of age, as already pointed out by Elliott (1979), who observed significant differences between 13-year-olds and 17-year-olds in speech-in-babble-noise perception. Indeed, speech-in-babble-noise perception depends on a wide range of cognitive abilities, whose development extend well beyond 11 years of age, such as sustained attention (Betts et al., 2006; Hoyer et al., 2021; Thillay et al., 2015), distractibility (Hoyer et al., 2021; Wetzels et al., 2009), impulsivity and motor control (Booth et al., 2003; Wright et al., 2003), verbal short-term memory (Alloway et al., 2006; Gathercole et al., 2004), and the ability to use pitch cues to focus on the target, including pitch short-

term memory (Ginzburg et al., 2022). This explains the great difference in the amount of noise that children and young adults can sustain to achieve the same performance, with more than 3 dB of difference in SNR between 10-year-olds and young adults, and more than 9 dB between 5-year-olds and adults in the difficult condition (see Figure 3). The latter result is very close to the 10 dB threshold difference reported between 5-year-olds and young adults, in a 2-talker masker, in Buss et al. (2019). This emphasizes the relevance of using a babble-noise in a speech perception screening test for auditory processing disorders, that is often associated with several deficits, such as attention deficits or short-term memory impairments, as speech perception performance in babble noise would be more sensitive to all those deficits than perception in speech-shaped noise.

Response times reveals competition processes within the lexicon

Response times reflected the expected slower speed of processing from the youngest to the eldest children (with about 500 ms of difference in response times observed here), due to processing speed maturation (Kail, 1991), in addition with slower response times for the oldest children as compared to adults. In a similar 4AFC task, Rigler et al. (2015) obtained a 400 ms difference between 16-year-olds and 9-year-olds having to click on the picture of a target word amongst 4 pictures, revealing that processing speed, as in the present study, undergoes maturation well beyond 10 years of age.

One mechanism that could explain the counterintuitive results observed in the present study of longer response times for the easy condition as compared to the difficult condition (for both children and adults) is the organisation of mental lexicon. Using eye tracking with the visual world paradigm (introduced by Huettig et al., 2011), Rigler et al. (2015) showed more activation for competition words for 9-year-olds than for 16-year-olds. Indeed, according to

most models for spoken word recognition (e.g., Frauenfelder & Floccia, 1999, for a review), words are organised within the mental lexicon as clusters of phonologically similar words (e.g., the neighbourhood activation model, Luce & Pisoni, 1998) and words sharing the same onset sounds (e.g. the Cohort model Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989). As a word begins to be heard, several potential candidates are activated within the mental lexicon and the process of selecting the correct target involves competition within the several activated candidates (e.g., the trace model, Dahan et al., 2001; McClelland & Elman, 1986). Several studies have shown that words with numerous phonological neighbours are more difficult to perceive and take a longer time to process, especially when those phonological neighbours are of higher occurrence frequency than the target word, than words with a low number of phonological neighbours (Dirks et al., 2001; Kirk et al., 1995; Krull et al., 2010). We would thus expect a greater response time when more numerous candidates are activated. In the present task, candidates are not only activated by the word heard, but, as well, by the three other words displayed (i.e., the target and the three foils). Here, in the difficult condition, the words are phonologically close, so the phonological patterns that are activated by the target words and the foils are similar, and are less numerous than the phonological patterns activated in the easy condition. Indeed, because the words, in the easy condition, are phonologically dissimilar, they would activate four different clusters of phonological patterns, demanding a greater number of comparisons between the target word and all the phonological similar words to the foils. The greater number of activated candidates could account for the longer response times in the ‘easy’ condition, even if this condition shows a higher accuracy.

Central hearing screening test design: impact of reproducibility, word familiarity, and psycholinguistic factors

Another major result obtained here is the absence of short-term practice effects. Indeed, by using a fixed SNR ratio method, rather than an adaptative one, we were able to assess the potential changes from the beginning of a block (12 first sets) and the end (12 last sets). Indeed, although we avoided the major source of practice effect by not using trial per trial visual feedback (Munro & Lutman, 2005), there are still a wide range of reasons why substantial modifications could occur. Firstly, the same sets of 4 pictures are used throughout, so more familiarity with the specific phonetic contrasts to listen to would be expected as the sets repeat themselves, especially in the difficult condition. Secondly, as the same feminine voice for the target words is used throughout, better identification and separation of the talker's voice from the babble background could be expected with practice. Those factors could account for a significant improvement of the accuracy with time. Conversely, a deterioration of accuracy could account for lack of maintenance of the necessary listening effort, involving a strong focused attention component, during the entire 2-minute-long blocks, for such a challenging task. We did not observe any significant difference in accuracy between series, whether the whole population was considered, or only the youngest ones (i.e., first quartile, < 7.3 years old). This shows the absence of learning during the course of one block with the task used here, and, as well, the absence of loss of motivation during the course of a block, which is particularly important when addressing young children. Indeed, to maintain motivation, Fallon et al., (2000) used a "game-like" automated visual feedback for correct and incorrect responses in a 4AFC in babble noise, and obtained a significant improvement between the first 20 trials (82%) and the last 20 trials (86%) of a 40 words block at constant SNR, in their children and adult groups.

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

In order to be able to use the test later in an adaptive manner, a stability of scores, at least within one block, is recommended.

Performance is likely to be influenced by familiarity with the target words depending on children's age (Brown & Watson, 1987), as contextual influences on those words develop. To assess this possibility, Wilson et al. (2010) checked the performance of their youngest groups in quiet, and obtained significant lower performance in the 6-year-olds versus the older children. They concluded that part of the 6-year-olds increased threshold was probably due to less familiarity for the target words. Although the pictures we used were designed so that 5-year-olds could recognize them easily, a growing familiarity as age increases is difficult to rule out. To address this question, we ranked each target word according to its score and correlated the ranking in each age group with the overall ranking, as in Fallon et al. (2000), who used a 4AFC task as well. We obtained quite similar results (r ranging from .83 to .92) as in their study ($r > .90$), showing that the relative difficulty of each target word didn't depend on age, and that the improvement of accuracy observed with age was not likely due to an increase in familiarity of the target words with age.

Although the experimental design was not aimed at testing specifically the influence of psycholinguistic factors on the test accuracy, it is important to ascertain whether they have substantial influence on the accuracy at a test aimed at screening for speech in babble noise perception deficits. Indeed, the influence of such psycholinguistic parameters is rarely considered when developing auditory perception tests. Yet, those parameters can introduce unwanted variability in audiological speech perception accuracy (Moulin et al., 2017), especially in interaction with cognitive status and lexical knowledge (Moulin & Richard, 2015). In a 4AFC task specifically designed to test the potential influence of occurrence frequency, by comparing scores for the same

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

target words presented with foils of high and low occurrence frequencies, a word frequency effect was observed and was significantly greater in adults than in 3- to 7-year-olds (Elliott et al., 1983). Furthermore, this frequency effect, in children, increased significantly with receptive vocabulary knowledge, which argues for the need to control for that effect when developing audiological tests of speech perception (Elliott et al., 1983). In the present test, no effect of target-word occurrence frequency was found, nor any effect of phonological neighbourhood density. This is important to consider for further refinement in the choice of target words, to ensure that scores reflect more auditory perception (Kosky & Boothroyd, 2003; Mendel, 2008) than mental lexicon development.

Overall, this study showed the neurodevelopmental course of speech perception in a babble noise that combines both energetic and informational masking, in primary school children, and confirms that processes involved in speech perception in babble noise (in particular informational masking) show a definite maturation step around 7 years of age and that they are far from mature at 10 years of age. We also unveiled the effect of lexicon competition processes on speed processing and checked for crucial factors that can influence such central hearing screening tests, such as word occurrence frequency and phonological neighborhood density. This resulted in a screening test that is quick, reproducible, engaging for children, devoid of short-term learning effect and that has therefore good potential for screening speech perception in babble noise deficits.

References

- Ahmed, A. U. (2021). Auditory Processing Disorder and Its Comorbidities : A Need for Consistency in Test Cutoff Scores. *American Journal of Audiology*, *30*(1), 128-144. https://doi.org/10.1044/2020_AJA-20-00103
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children : Are They Separable? *Child Development*, *77*(6), 1698-1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- Anderson, K. (2004). The Problem of Classroom Acoustics : The Typical Classroom Soundscape Is a Barrier to Learning. *Seminars in Hearing*, *25*(02), 117-129. <https://doi.org/10.1055/s-2004-828663>
- Bamiou, D., Musiek, F., & Luxon, L. (2001). Aetiology and clinical presentations of auditory processing disorders—A review. *Archives of Disease in Childhood*, *85*(5), 361-365. <https://doi.org/10.1136/ad.85.5.361>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. <https://doi.org/10.48550/ARXIV.1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.
- Betts, J., McKay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children : The effect of age and task load. *Child Neuropsychology*, *12*(3), 205-221.
- Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2003). Neural development of selective attention and response inhibition. *Neuroimage*, *20*(2), 737-751.
- Bourgeois-Vionnet, J., Moulin, A., Hermier, M., Pralus, A., & Nighoghossian, N. (2020). A case of verbal and emotional prosody processing dissociation after a right temporal venous infarct. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *41*(6), 1615-1618. <https://doi.org/10.1007/s10072-019-04175-w>
- Bradley, J. S., & Sato, H. (2008). The intelligibility of speech in elementary school classrooms. *The Journal of the Acoustical Society of America*, *123*(4), 2078-2086. <https://doi.org/10.1121/1.2839285>
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). *Speaking Clearly for Children With Learning Disabilities Sentence Perception in Noise*. 18.
- Bregman, A. S. (1994). *Auditory scene analysis : The perceptual organization of sound*. MIT press.

- Brown, G. D., & Watson, F. L. (1987). First in, first out : Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & cognition*, *15*, 208-216.
- Buss, E., Felder, J., Miller, M. K., Leibold, L. J., & Calandruccio, L. (2022). Can Closed-Set Word Recognition Differentially Assess Vowel and Consonant Perception for School-Age Children With and Without Hearing Loss? *Journal of Speech, Language, and Hearing Research*, *65*(10), 3934-3950. https://doi.org/10.1044/2022_JSLHR-20-00749
- Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked Sentence Recognition in Children, Young Adults, and Older Adults : Age-Dependent Effects of Semantic Context and Masker Type. *Ear and Hearing*, *40*(5), 1117-1126. <https://doi.org/10.1097/AUD.0000000000000692>
- Buss, E., Leibold, L. J., & Hall, J. W. (2016). Effect of response context and masker type on word recognition in school-age children and adults. *The Journal of the Acoustical Society of America*, *140*(2), 968-977. <https://doi.org/10.1121/1.4960587>
- Cabbage, K. L., & Hitchcock, E. R. (2022). Clinical Considerations for Speech Perception in School-Age Children With Speech Sound Disorders : A Review of the Current Literature. *Language, Speech & Hearing Services in Schools*, *53*(3), 768-785. https://doi.org/10.1044/2022_LSHSS-21-00120
- Calandruccio, L., Leibold, L. J., & Buss, E. (2016). Linguistic Masking Release in School-Age Children and Adults. *American Journal of Audiology*, *25*(1), 34-40. https://doi.org/10.1044/2015_AJA-15-0053
- Calandruccio, L., & Zhou, H. (2014). Increase in Speech Recognition due to Linguistic Mismatch Between Target and Masker Speech : Monolingual and Simultaneous Bilingual Performance. *Journal of Speech Language and Hearing Research*, 1089. https://doi.org/10.1044/2013_JSLHR-H-12-0378
- Calcus, A., Deltenre, P., Colin, C., & Kolinsky, R. (2018). Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Developmental Science*, *21*(3), e12558. <https://doi.org/10.1111/desc.12558>
- Carvalho, N. G. de, Novelli, C. V. L., & Colella-Santos, M. F. (2017). Evaluation of speech in noise abilities in school children. *International Journal of Pediatric Otorhinolaryngology*, *99*, 66-72. <https://doi.org/10.1016/j.ijporl.2017.05.019>
- Coffeen, B. (2000). *Classroom Acoustics*. Technical committee of architectural acoustics of the acoustical society of america. <https://acousticalsociety.org/classroom-acoustics-booklets/>
- Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of Open-Set Word Recognition in Children : Speech-Shaped Noise and Two-Talker Speech Maskers. *Ear and hearing*, *37*(1), 55-63.
- Crandell, C. C., & Smaldino, J. J. (2000). Classroom Acoustics for Children With Normal Hearing and With Hearing Impairment. *Language, Speech, and Hearing Services in Schools*, *31*(4), 362-370. <https://doi.org/10.1044/0161-1461.3104.362>

- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534. <https://doi.org/10.1080/01690960143000074>
- Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., & Fausti, S. A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing*, *22*(1), 1-13.
- Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America*, *66*(3), 651-653. <https://doi.org/10.1121/1.383691>
- Elliott, L. L., Clifton, L. A., & Servi, D. G. (1983). Word frequency effects for a closed-set word identification task. *Audiology*, *22*(3), 229-240. <https://doi.org/10.3109/00206098309072787>
- Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker babble. *The Journal of the Acoustical Society of America*, *108*(6), 3023-3029. <https://doi.org/10.1121/1.1323233>
- Finitzo, -Hieber Terese, & Tillman, T. W. (1978). Room Acoustics Effects on Monosyllabic Word Discrimination Ability for Normal and Hearing-Impaired Children. *Journal of Speech and Hearing Research*, *21*(3), 440-458. <https://doi.org/10.1044/jshr.2103.440>
- Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.
- Frauenfelder, U. H., & Floccia, C. (1999). The Recognition of Spoken Words. In A. D. Friederici, *Language Comprehension: A Biological Perspective* (p. 1-40). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59967-5_1
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5), 2246-2256.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177-190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5- to 10-year-old children: Focus on musical and verbal short-term memory. *Developmental Science*, *25*(3). <https://doi.org/10.1111/desc.13188>
- Guzek, A., & Iwanicka-Pronicka, K. (2022). Analysis of the auditory processing skills in 1,012 children aged 6-9 confirms the adequacy of APD testing in 6-year-olds. *PLOS ONE*, *17*(8), e0272723. <https://doi.org/10.1371/journal.pone.0272723>
- Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children.

- Ear and Hearing*, 23(2), 159-165. <https://doi.org/10.1097/00003446-200204000-00008>
- Hétu, R., Truchon-Gagnon, C., & Bilodeau, S. A. (1990). Problems of noise in school settings : A review of literature and the results of an exploratory study. *Journal of Speech-Language Pathology and Audiology*, 14, 31-39.
- Hillock-Dunn, A., Taylor, C., Buss, E., & Leibold, L. J. (2015). Assessing Speech Perception in Children With Hearing Loss : What Conventional Clinical Tools May Miss. *Ear and hearing*, 36(2), e57-e60.
- Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research: JSLHR*, 41(1), 94-106. <https://doi.org/10.1044/jslhr.4101.94>
- Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology*, 49(12), 928-932. <https://doi.org/10.3109/14992027.2010.520036>
- Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *Journal of Abnormal Psychology*, 54(1), 75-85.
- Hoyer, R. S., Elshafei, H., Hemmerlin, J., Bouet, R., & Bidet-Caulet, A. (2021). Why Are Children So Distractible? Development of Attention and Motor Control From Childhood to Adulthood. *Child Development*, 92(4). <https://doi.org/10.1111/cdev.13561>
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing : A review and critical evaluation. *Acta Psychologica*, 137(2), 151-171. <https://doi.org/10.1016/j.actpsy.2010.11.003>
- Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, 109(3), 490-501. <https://doi.org/10.1037/0033-2909.109.3.490>
- Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 40(6), 1395-1405.
- Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*, 16(5), 470-481.
- Koopmans, W. J. A., Goverts, S. T., & Smits, C. (2018). Speech Recognition Abilities in Normal-Hearing Children 4 to 12 Years of Age in Stationary and Interrupted Noise: *Ear and Hearing*, 39(6), 1091-1103. <https://doi.org/10.1097/AUD.0000000000000569>
- Kosky, C., & Boothroyd, A. (2003). Validation of an On-line Implementation of the Imitative Test of Speech Pattern Contrast Perception (IMSPAC). *Journal of the American Academy of Audiology*, 14(02), 072-083. <https://doi.org/10.3766/jaaa.14.2.3>

- Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical effects on spoken-word recognition in children with normal hearing. *Ear and Hearing, 31*(1), 102-114. <https://doi.org/10.1097/AUD.0b013e3181b7892f>
- Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://CRAN.R-project.org/package=emmeans>
- Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in congenital amusia: Impaired recognition, but preserved emotional intensity. *Neuropsychology, 32*(7), 880-894. <https://doi.org/10.1037/neu0000461>
- Lewis, H. D., Benignus, V. A., Muller, K. E., Malott, C. M., & Barton, C. N. (1988). Babble and Random-Noise Masking of Speech in High and Low Context Cue Conditions. *Journal of Speech, Language, and Hearing Research, 31*(1), 108-114. <https://doi.org/10.1044/jshr.3101.108>
- Linares, D., & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal, 8*(1), 122-131. <https://doi.org/10.32614/RJ-2016-008>
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing, 19*(1), 1.
- Lüdtke, D. (2019). *esc: Effect Size Computation for Meta Analysis (Version 0.5.1)*. <https://doi.org/10.5281/zenodo.1249218>
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia: The Importance of Serial Order. *Frontiers in Psychology, 7*. <https://doi.org/10.3389/fpsyg.2016.01522>
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.
- Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance, 15*(3), 576.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.
- McMillan, B. T. M., & Saffran, J. R. (2016). Learning in Complex Environments: The Effects of Background Speech on Early Word Learning. *Child Development, 87*(6), 1841-1855. <https://doi.org/10.1111/cdev.12559>
- Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology, 47*(9), 546-553. <https://doi.org/10.1080/14992020802252261>
- Meyer, T. A., & Pisoni, D. B. (1999). Some computational analyses of the PBK test: Effects of frequency and lexical density on spoken word recognition. *Ear and Hearing, 20*(4), 363-371.
- Moore, D. R. (2018). Editorial: Auditory Processing Disorder. *Ear and Hearing, 39*(4), 617-620. <https://doi.org/10.1097/AUD.0000000000000582>

- Moulin, A., Bernard, A., Tordella, L., Vergne, J., Gisbert, A., Martin, C., & Richard, C. (2017). Variability of word discrimination scores in clinical practice and consequences on their sensitivity to hearing loss. *European Archives of Oto-Rhino-Laryngology: Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS): Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery*, 274(5), 2117-2124. <https://doi.org/10.1007/s00405-016-4439-x>
- Moulin, A., Favier, C., Richard, C., & Farget, V. (2013). French Babble Noises for Speech Perception Tests. Development and Spectro-temporal analysis. *Annales D'oto-Laryngologie Et De Chirurgie Cervico Faciale*.
- Moulin, A., & Richard, C. (2015). Lexical Influences on Spoken Spondaic Word Recognition in Hearing-Impaired Patients. *Auditory Cognitive Neuroscience*, 476. <https://doi.org/10.3389/fnins.2015.00476>
- Munro, K. J., & Lutman, M. E. (2005). The influence of visual feedback on closed-set word test performance over time. *International Journal of Audiology*, 44(12), 701-705. <https://doi.org/10.1080/14992020500271795>
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661-677. <https://doi.org/10.1017/S014271640707035X>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516-524. <https://doi.org/10.3758/BF03195598>
- Plaza, M., Cohen, H., & Chevrie-Muller, C. (2002). Oral Language Deficits in Dyslexic Children: Weaknesses in Working Memory and Verbal Planning. *Brain and Cognition*, 48(2-3), 505-512. <https://doi.org/10.1006/brcg.2001.1407>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, 51, 1690-1703. <https://doi.org/10.1037/dev0000044>
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431-2443. <https://doi.org/10.1121/1.4794379>
- Sato, H., & Bradley, J. S. (2008). Evaluation of acoustical conditions for speech communication in working elementary school classrooms. *The Journal of the Acoustical Society of America*, 123(4), 2064-2077.
- Savin, H. B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America*, 35(2), 200-206. <https://doi.org/10.1121/1.1918432>

- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, *118*(5), 2775-2778. <https://doi.org/10.1121/1.2062650>
- Skarżyński, H., & Piotrowska, A. (2012). Screening for pre-school and school-age hearing problems : European Consensus Statement. *International Journal of Pediatric Otorhinolaryngology*, *76*(1), 120-121. <https://doi.org/10.1016/j.ijporl.2011.10.016>
- Skarzynski, P. H., Wlodarczyk, A. W., Kochanek, K., Pilka, A., Jedrzejczak, W., Olszewski, L., Bruski, L., Niedzielski, A., & Skarzynski, H. (2015). Central auditory processing disorder (CAPD) tests in a school-age hearing screening programme – analysis of 76,429 children. *Annals of Agricultural and Environmental Medicine*, *22*(1), 90-95. <https://doi.org/10.5604/12321966.1141375>
- Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*, *18*(2), 89.
- Thillay, A., Roux, S., Gissot, V., Carreau-Martin, I., Knight, R. T., Bonnet-Brilhault, F., & Bidet-Caulet, A. (2015). Sustained attention and prediction : Distinct brain maturation trajectories during adolescence. *Frontiers in Human Neuroscience*, *9*, 519.
- Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia : Evidence from a gating paradigm. *Cortex*, *59*, 84-94. <https://doi.org/10.1016/j.cortex.2014.07.012>
- Vickers, D. A., Moore, B. C. J., Majeed, A., Stephenson, N., Alferaih, H., Baer, T., & Marriage, J. E. (2018). Closed-Set Speech Discrimination Tests for Assessing Young Children. *Ear & Hearing*, *39*(1), 32-41. <https://doi.org/10.1097/AUD.0000000000000528>
- Wetzel, N., Widmann, A., & Schröger, E. (2009). The cognitive control of distraction by novelty in children aged 7–8 and adults. *Psychophysiology*, *46*(3), 607-616.
- Wightman, F., & Allen, P. (1992). Individual differences in auditory capability among preschool children. In *Developmental psychoacoustics* (p. 113-133). American Psychological Association. <https://doi.org/10.1037/10119-004>
- Wilson, R. H., Farmer, N. M., Gandhi, A., Shelburne, E., & Weaver, J. (2010). Normative Data for the Words-in-Noise Test for 6- to 12-Year-Old Children. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1111-1121. [https://doi.org/10.1044/1092-4388\(2010/09-0270\)](https://doi.org/10.1044/1092-4388(2010/09-0270))
- Wilson, W. J., & Arnott, W. (2013). Using Different Criteria to Diagnose (Central) Auditory Processing Disorder : How Big a Difference Does It Make? *Journal of Speech, Language, and Hearing Research*, *56*(1), 63-70. [https://doi.org/10.1044/1092-4388\(2012/11-0352\)](https://doi.org/10.1044/1092-4388(2012/11-0352))
- Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development : Typical

developmental trends. *Journal of Child Psychology and Psychiatry*, 44(4), 561-575.

Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear and hearing*, 33(6), 731-744.

<https://doi.org/10.1097/AUD.0b013e31825aeced>

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005).

Deficits in speech perception predict language learning impairment.

Proceedings of the National Academy of Sciences, 102(39), 14110-14115.

<https://doi.org/10.1073/pnas.0504446102>

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12(5), 732-745.

<https://doi.org/10.1111/j.1467-7687.2009.00817.x>

Supplementary material

Written form	Phonological form	Frequency of occurrence (per million occurrences)
bain	bɛ̃	50.52
banc	bɑ̃	8.96
beurre	bœʀ	15.12
blanc	blɑ̃	53.93
chapeau	ʃapo	48.61
chateau	ʃato	40.51
coeur	kœʀ	224.98
cou	ku	43.71
coule	kul	14.7
dent	dɑ̃	13.27
doigt	dwa	39.83
drapeau	drapo	14.66
fleur	flœʀ	25.2
gant	gɑ̃	9.86
heure	œʀ	415.4
main	mɛ̃	286.62
nain	nɛ̃	9.08
pain	pɛ̃	62.81
rateau	rato	0.77
roi	rwa	166.34
roue	ru	13.49
rouge	ruʒ	79.7
toit	twa	42.63
trois	trwa	380.8

Table S1 : All target words: their written form, their phonological form, and their frequency of occurrence in the French language

Section 2: Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

Difficulty condition	Phonological form of every 4-item sets	Average Levenshtein distance within a set	Number of unique phonological neighbors for each set
Phonologically distant (easy condition)	bã / floer / mē / trwa	3.67	76
	bē / koer / kul / rato	3.33	86
	boer / dā / dwa / pē	2.67	95
	blã / ſato / ku / twa	3.5	86
	Drapo / gã / oer / ru	3.17	80
	ſapo / nē / rwa / ru3	3.33	77
Phonologically close (difficult condition)	bã / blã / dã / gã	1.33	73
	bē / mē / nē / pē	1	62
	boer / koer / floer / oer	1.5	51
	dwa / rwa / twa / trwa	1.17	45
	ku / kul / ru / ru3	1.5	83
	ſapo / ſato / drapo / rato	1.83	44

Table S2 : All four-word sets used during the speech-in-noise task for the Easy condition and for the Difficult condition. For each set we report the average Levenshtein distance between all pairs of items within the set and the number of unique phonological neighbors.

3 Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

Article published as: Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5 - to 10 - year - old children: Focus on musical and verbal short - term memory. *Developmental Science*, 25(3), e13188. <http://doi.org/10.1111/desc.13188>

Available at <https://hal.science/hal-03434135v1>

3.1 General introduction

Auditory short-term memory (STM) is a fundamental cognitive function that plays a pivotal role in child development. It enables individuals to temporarily hold in mind auditory information, such as spoken language and musical tones for brief periods of time. Auditory STM contributes to various cognitive processes critical for effective communication. As children progress through different stages of their development, from early childhood to adolescence, the maturation of auditory STM is intricately connected to the refinement of their language and reading abilities. In the present study, we aimed at characterizing the developmental trajectory of auditory STM in school-aged children for two types of auditory materials: musical and verbal. We used a child-adapted delayed-matching-to-sample task (DMST) that allowed the direct comparison of auditory STM for both materials, enabling us to unveil common and distinct mechanisms for the processing of the two materials in auditory STM. Moreover, we explored its relation to children's speech in babble-noise perception, in order to disclose the common mechanisms between these two fundamental auditory processes.

3.2 Article

Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

Jérémie Ginzburg^{1,2}, Annie Moulin^{1,2}, Lesly Fornoni^{1,2}, Francesca Talamini³,
Barbara Tillmann^{1,2}, Anne Caclin^{1,2}

¹ Lyon Neuroscience Research Center; CNRS, UMR5292; INSERM, U1028; Lyon, F-69000, France

² University Lyon 1, Lyon, F-69000, France

³ Institute für Psychologie, Universität Innsbruck, Austria

Running title: Auditory short-term memory development

Conflict of interest statement: the authors declare no conflict of interest.

Data availability statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Acknowledgments: We thank all the children who took part in the study, along with their parents, and their welcoming teachers. We also thank Nathalie Bedoin and Emmanuel Ferragne for their help for the creation of the stimuli, as well as Céline Hidalgo and Daniele Schön for their help with “p-center” calculation.

Funding: This work was conducted within the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) and of the LabEx Cortex (“Construction, Cognitive Function and Rehabilitation of the Cortex”, ANR-11-LABX-0042) of Université de Lyon, within the program “Investissements d’avenir” (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). This work was funded by a Pack Ambition Recherche (COGAUDYS project) from the Auvergne-Rhône-Alpes Region to AC, AM, and BT.

Research Highlights

Auditory short-term memory was assessed in 5- to 10-year-old children and young adults using musical and verbal materials in a delayed matching-to-sample task.

Musical and verbal short-term memory shared a similar developmental trajectory and are still under development at 10 years of age.

Correlations with speech perception in cocktail-party noise suggest shared cognitive resources between musical short-term-memory and speech in cocktail-party noise perception capacities.

Testing both musical and verbal short-term memory provides perspectives for diagnosis and training in developmental learning disorders.

Abstract

Developmental aspects of auditory cognition were investigated in 5-to-10-year-old children ($n = 100$). Musical and verbal short-term memory (STM) were assessed by means of delayed matching-to-sample tasks (comparison of two 4-item sequences separated by a silent retention delay), with two levels of difficulty. For musical and verbal materials, children's performance increased from 5 years to about 7 years of age, then remained stable up to 10 years of age, with performance remaining inferior to performance of young adults. Children and adults performed better with verbal material than with musical material. To investigate auditory cognition beyond STM, we assessed speech-in-noise perception with a 4-alternative forced-choice task with two conditions of phonological difficulty and two levels of cocktail-party noise intensity. Partial correlations, factoring out the effect of age, showed a significant link between musical STM and speech-in-noise perception in the condition with increased

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

noise intensity. Our findings reveal that auditory STM improves over development with a critical phase around 6-7 years of age, yet these abilities appear to be still immature at 10 years. Musical and verbal STM might in particular share procedural and serial order processes. Furthermore, musical STM and the ability to perceive relevant speech signals in cocktail-party noise might rely on shared cognitive resources, possibly related to pitch encoding. To the best of our knowledge, this is the first time that auditory STM is assessed with the same paradigm for musical and verbal material during childhood, providing perspectives regarding diagnosis and remediation in developmental learning disorders.

Keywords: working memory, auditory perception, music, speech, recognition task, speech-in-noise

Introduction

Auditory Short-Term Memory (STM) allows for encoding, storage, and retrieval of auditory information during a short amount of time (within several seconds, Baddeley & Hitch, 1974; Cowan, 2008)¹. Along with auditory scene analysis and auditory attention, it is a key component of central auditory processing that subtends auditory cognition and allows making sense of the ever-changing acoustic environment. Numerous relationships have been described between central auditory processing disorders (CAPD) and learning disabilities in children (Goswami, 2011; Iliadou & Iakovides, 2003; Medwetsky, 2011; Moore et al., 2010). Recent research suggests a link between verbal STM deficits and learning disorders (Männel et al., 2015; Nithart et al., 2009; Perez et al., 2012) and a link between impaired speech-in-noise processing and learning disorders (Bradlow et al., 2003; Sperling et al., 2005; Ziegler et al., 2009, 2011). Specifically, verbal STM impairment has been observed in dyslexic children with reduced digit spans or poor non-word repetition in recall tasks (Adlard & Hazan, 1998; Kramer et al., 2000; Majerus & Cowan, 2016; Nithart et al., 2009; Plaza et al., 2002; Roodenrys & Stokes, 2001; Tijms, 2004). The link between impaired verbal STM or working memory (WM) and learning disorders has also been reported for specific language impairment (SLI, Nithart et al., 2009) and dyscalculia (Attout & Majerus, 2015). It appears that both verbal STM and speech-in-noise perception are consistently reported to be impaired in learning disorders and CAPD (Moore et al., 2010; Perez et al., 2012; Ziegler et al., 2011).

Similarly to verbal STM deficits observed in several language-related disorders (Adlard & Hazan, 1998; Kramer et al., 2000; Majerus & Cowan, 2016;

¹ Note that “Working memory” is sometimes used in a larger sense encompassing this STM definition. Here we will refer to working memory only when the paradigm entails a manipulation of information.

Nithart et al., 2009; Plaza et al., 2002; Roodenrys & Stokes, 2001; Tijms, 2004), musical STM deficits have also been reported in children with neurodevelopmental learning disorders. Dyslexic children display lower performance than typically-developing children in recognition STM tasks for pitch (Ziegler et al., 2012) and in tonal recognition tasks from the Primary Measure of Music Audiation (PMMA, Gordon, 1986) as shown by Atterbury (1985) and Forgeard et al. (2008). Furthermore, a sizeable comorbidity between dyslexia and congenital amusia has been observed in adults and children (Couvignou et al., 2019; Couvignou & Kolinsky, 2021). Congenital amusia is characterized by a deficit in music processing and in particular musical STM (Tillmann et al., 2009, 2016). These findings stress the importance of characterizing the development of central auditory processing in typically developing children, in particular to improve diagnosis and rehabilitation of central auditory processing deficits associated with learning disorders.

It has been well documented that verbal short-term storage capacity increases during childhood (Alloway et al., 2006; Chuah & Maybery, 1999; Cowan et al., 1999; Dempster, 1981; Gathercole, 1999; Gathercole et al., 2004; Orsini et al., 1987). Gathercole (1999) considers age as the most powerful factor influencing verbal STM capacity. These verbal STM abilities, tested with digit, word, and non-word spans (i.e., recall tasks), increase linearly with age between 4 to 14 years of age and appear to level-off by the age of 14-15 years (Gathercole et al., 2004). The development of STM over childhood was thus mostly investigated using serial recall paradigms that are reliant on verbal production, hence specific to verbal and/or phonological material (Gathercole, 1999). Only a few studies have investigated children's STM for other types of auditory information (e.g., pitch, timbre or rhythm and temporal processing), but not yet over development nor in comparison to verbal material. Reviewing previous

research reveals that only few studies investigated musical STM over children development. Pitch memory for single tones arise as early as 6 months of age (Plantinga & Trainor, 2008). Later in development, STM for tone sequences appears to mature from early childhood (six years old) to pre-teenage years (thirteen years old, Clark et al., 2018). Indeed, Clark et al. (2018) found a developmental increase in the memory capacity for single tones as well as tone sequences similar to the developmental trajectory of visuo-spatial memory in children from 6- to 13-years of age, using an adapted part-set cueing task. In addition to an increase of capacity, Keller and Cowan (1994) found a decrease in STM trace decay for pitch over time in children from 4- to-12 years old using a 2-tone comparison task. This latter study found an increase in the persistence of memory for pitch between the ages of 6 to 7 years. To our knowledge, no study has compared directly the precise development of musical and verbal STM. Our study aimed at doing so systematically in children from 5- to 10-years old by using a paradigm that allows for direct comparison between musical and verbal STM.

Musical and verbal STM are difficult to compare if standard recall paradigms (most frequently used to evaluate verbal STM) are used. Even if some studies compared musical and verbal STM using mixed recall/reconstruction paradigms (Gorin et al., 2016, 2018; Williamson et al., 2010), they nonetheless required production processes that are difficult to adapt for children. The classical delayed matching-to-sample task (DMST) allows us to circumvent the need of oral or motor production that is required in a recall paradigm. In a DMST, participants have to memorize a first (S1) sequence of sounds (or an isolated sound). After a delay, a second stimulus is presented (S2) and participants have to report whether S1 and S2 are identical or different. This task has the advantage of allowing for the use of different kinds

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

of materials (verbal, musical, environmental sound..., e.g., Talamini et al., 2021) and entails the three memorization steps of encoding (during S1), retention (during the delay), and retrieval (during S2) without relying on a production phase (as required in recall tasks). Hence, the DMST appears to be a well-suited paradigm to assess the development of auditory STM (for both musical and verbal materials) in children.

In adults, it has been suggested that auditory STM, rather than being a unitary phenomenon, could be based on partly separate subsystems for different types of material, in particular musical and verbal information (for a review, see Caclin & Tillmann, 2018). Berz (1995) and Pechmann & Mohr (1992) proposed a musical/tonal loop that would account for a storage component specific to the representational features of tonal information, based on Baddeley & Hitch (1974) multicomponent model of WM. Ockelford (2007), based on the same multicomponent model, has also suggested to add a musical central executive component that would entail attentional processes specific to musical information. These models might predict different maturation patterns for each material type, leading to domain-specific patterns of developmental trajectory (i.e. different patterns of recency effects and absence of correlations between musical and verbal STM). In another line of research, other models have postulated more general attentional processes involved in the maintaining of information in a short-term storage and more specific item-related processes concerning the encoding of information (Barrouillet & Camos, 2007; Cowan, 1998). These latter models would predict a greater involvement of domain-general attentional processes in STM for both materials and would probably predict a similar developmental trajectory for musical and verbal STM as well as similitudes of domain-general processes between them (similar recency effects and correlations between both materials).

In addition, to further our understanding of the distinct mechanisms between musical and verbal STM, our study aimed at testing the development of auditory cognition beyond STM. It has been proposed that STM and speech-in-noise perception might share encoding processes (Murphy et al., 2000; Pichora-Fuller et al., 1995; Sarampalis et al., 2009), and both are deficient in learning disorders (Perez et al., 2012; Ziegler et al., 2009). In order to confirm that STM and speech-in-noise perception share encoding processes and to shed light on material-specific perceptual processes shared by auditory STM and speech-in-noise, we compared musical and verbal STM with speech-in-noise perception abilities. The two types of WM models described in the previous paragraph would both predict different encoding processes for musical and verbal material. Consequently, if STM and speech-in-noise share encoding processes, we predict differential links between musical and verbal STM, and speech-in-noise (e.g. only musical STM is linked to speech-in-noise perception or only verbal STM). If only musical STM shares processes with speech-in-noise, these results would be in line with the already observed reliance of sound segregation on pitch processing (Oxenham, 2008) and bring evidence for the observed musician-advantage for speech-in-noise processing (Chandrasekaran & Kraus, 2010; Parbery-Clark et al., 2009; Strait et al., 2012; Zendel et al., 2015). If, however, only verbal STM is linked to speech-in-noise, the evidence would be in favor of top-down lexical influences or fine temporal structure encoding shared by speech-in-noise perception and verbal STM (Zekveld et al., 2013).

The aims of the present work are thus threefold. First, this study aims at describing the developmental trajectory of musical and verbal STM as, until now, only the trajectory of verbal STM using recall tasks has been described. The use of a DMST provides the potential to investigate specific patterns of development that could be task-related and/or material-specific, as it allows us

to directly compare musical and verbal STM. Second, the present study aims at bringing insights about the shared and distinct mechanisms between musical and verbal STM, as domain-general and domain-specific processes have been described for musical and verbal STM in adults (Gorin et al., 2016, 2018). Third, in order to go beyond auditory STM and to scrutinize domain-specific processes in auditory STM, dependences between auditory STM and speech-in-noise perception were explored using cocktail party noise. As speech-in-noise and auditory STM would share encoding processes (Murphy et al., 2000; Pichora-Fuller et al., 1995; Sarampalis et al., 2009), finding out if both materials are linked to speech-in-noise or only one of them would bring insights into the existence of shared (e.g. both materials linked to speech-in-noise) or distinct (e.g. one of the materials linked to speech-in-noise) encoding processes.

In the present study, we created a child-adapted DMST with 4-tone sequences for the musical material and with sequences of four consonant-vowel syllables for the verbal material. We implemented two levels of difficulty and tested 100 children ranging from 5 to 10 years of age (from kindergarten to 5th grade). The same 100 children were tested with a 4-alternative forced-choice speech-in-noise perception task.

Methods

Participants

One hundred children (mean age = 7.6 y.o ; min = 54 month-old, max = 127 month-old, 6 left-handed), attending a public primary school in South-East of France, participated in this study. The children were tested during school hours. Participation in the study was proposed to a total of 114 children and they were included in the study only if both parents or legal tutors provided written informed consent. The study was approved by the relevant services of the French public education services (Inspection de l'Education Nationale

(IEN) and Direction Académique des Services de l'Éducation Nationale (DASEN) of the Isere department). Children from Kindergarten (n=12), 1st (n=17), 2nd (n=18), 3rd (n=19), 4th (n=16), and 5th (n=18) grade underwent the experiment (Table 1). Among the 100 children, parents' responses to questionnaires revealed that 5 children had a diagnosed learning disability (dyslexia, dysphasia, dyscalculia, dysorthographie, or dysgraphie), 30 had seen at least once a speech-therapist and 6 had already worn grommets. Nine out of the hundred children had specific musical training for more than 2 months (mean = 2.01 years, SD = 1.29), other children had never had any musical training apart from normal school curriculum. As we aimed here to explore the cognitive abilities of children among a representative set of the population, we present the results including all children.

Twelve adults (mean age = 26.5 years, SD = 9.6, one left-handed) were also included in the study. None reported any neurological or psychiatric troubles. Six of them had had a few years of musical education (mean = 4 years, SD = 2.8) but none of them were practicing any instrument at the time of the experiment and this for at least the last 10 years. Their level of education, along with children's parental level of education, are shown in Figure 1.

English label	Kindergarten (KG)	1 st Grade	2 nd Grade	3 rd Grade	4 th Grade	5 th Grade
French label	Grande Section de Maternelle (GS)	Cours préparatoire (CP)	Cours élémentaire 1 (CE1)	Cours élémentaire 2 (CE2)	Cours Moyen 1 (CM1)	Cours Moyen 2 (CM2)
N	12	17	18	19	16	18
Mean age in years (SD)	5.01 (0.31)	6.03 (0.29)	6.93 (0.30)	8.01 (0.34)	8.92 (0.26)	9.94 (0.29)

Table 1 : Number of participants and mean age (SD in parentheses) for each grade and matching English and French label for educational level.

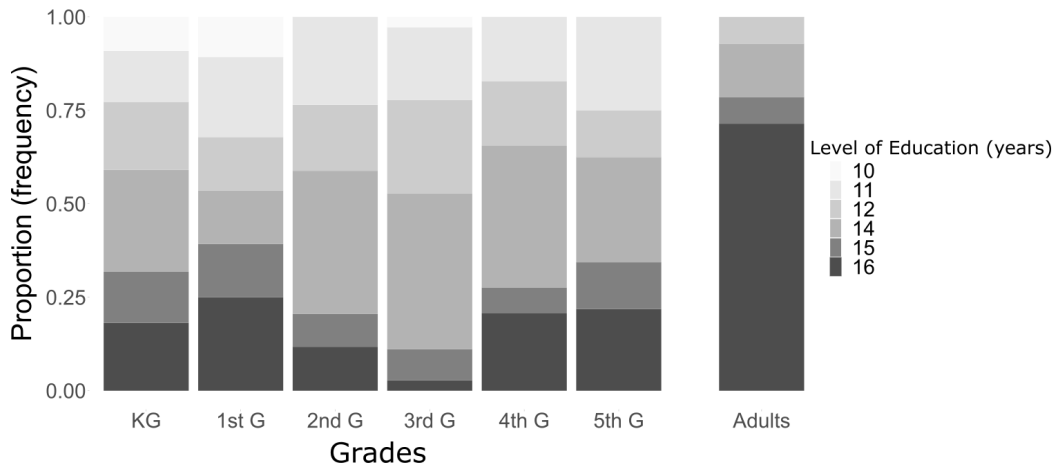


Figure 1 : Distribution of children’s parents and adult participants’ level of education. Level of education categories correspond to the number of years spent in scholar institutions, from first grade on. A Pearson’s Chi-squared test between all children’s grades and the number of parents per category of level of education revealed that the proportion of parental level of education did not significantly differ by grade $X^2(25, N = 181) = 25.97, p = .41$. A Chi-squared was also performed between adult’s distribution of level of education and the mean proportion of children’s parental level of education. The proportion of parental level of education differed by group (parents/adults) $X^2(5, N = 195) = 25.6, p = .00011$ ².

Stimuli construction and task design

Short-Term Memory task

As shown in Figure 2, each trial of the STM recognition task consisted in listening to a 4-item auditory sequence (S1), then after a silent retention delay of 2000 ms, to another 4-item sequence (S2) that could be identical or different. When S2 was different, a new item could appear equiprobably at the 2nd, 3rd, or 4th position of the sequence. Each item lasted 500 ms, the silent inter-stimulus interval (ISI) between two items lasted 100 ms so overall there was a 600 ms stimulus onset asynchrony (SOA), leading to a duration of 6600 ms for S1 and S2 sequences. Children were given unlimited time to give a response. The next trial started after a 1500 ms delay. Presentation® software (Version 18.0,

² We aimed at reaching homogeneity between children’s parental level of education and adult participants’ level of education but recruitment of adult participants fell during the Covid-19 pandemic, preventing us to fully achieve that goal. In the sample tested, all adult participants had very high level of performance irrespective of their education level

Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used to present stimuli and record responses in the STM task. Based on previous data-points in the literature (Gathercole et al. 2001; Jarrold, Thorn, et Stephens 2009; Majerus et al. 2006, 2007) we found that the use of 4-item sequences was an optimal trade-off for younger children to be able to do the task and for older children not being at ceiling.

There were two material conditions for the STM task: musical piano tones (music condition) or syllables (verbal condition).

For the musical STM task, six musical tones (piano timbre, Cubase database, Steinberg) belonging to the C major scale were used (C2, D2, E2, F2, G2, A2) with frequencies ranging from 131 to 220 Hz (thus encompassing the fundamental frequency range of the vowel recordings: 202-212 Hz). A total of 48 four-item sequences were generated, all tones were different within a given sequence and they all contained at least one ascending interval and one descending interval (to avoid simple, constantly rising or falling patterns that could facilitate memorization). Twenty-four S1 sequences were used as S2 sequences for the “same” trials, half for the easy condition and half for the difficult one. There was no difference between the difficulty conditions for the “same” trials. Twenty-four S2 sequences were generated for the “different” trials, which can be of two types depending on whether the change violated contour or not. Previous research has shown that a contour-violation leads to better performance in melody discrimination tasks than contour preservation (Dyson & Watkins, 1984; Monahan et al., 1987; Peretz & Babai, 1992; Ziegler et al., 2012). For the easy condition, twelve S2 sequences were generated with the new item in S2 changing the contour of the sequence (the contour is the up-and-down scheme of a melody,). So, if S1 had an up-down-up contour (e.g. E-A-D-F), S2 could have a down-up-up contour (e.g. E-C-D-F, Supplementary

Material, Sound 1). In the difficult condition, twelve S2 sequences were generated, the new item in S2 did not change the contour (Supplementary Materials, Sound 2). It should be noted that here, greater difficulty of the musical memory task is induced by not only the absence of contour change but also by smaller pitch differences between the new item and the original one in the difficult condition compared to the easy condition.

For the verbal STM task, the items were Consonant-Vowel syllables. To avoid difficulties due to children's phonological skills, the consonant-vowel stimuli were selected to show the greatest perceptual distance with each other, within a S1 (or S2) sequence Six consonants and six vowels were thus selected: /f/ /t/ /z/ /g/ /m/ /l/ and /i/ /e/ /a/ /y/ /ø/ /u/ resulting into 36 syllables that were then recorded by a professional mezzo-soprano singer (for details about syllables construction, see Supplementary Figure S1). A total of 48 four-item sequences were generated to be used as S1. All vowels and all consonants were different in a given sequence. Twenty-four S1 sequences were used as S2 sequences for the "same" trials, half for the easy condition and half for the difficult one. For the other twenty-four S1 sequences, twenty-four S2 sequences were created for the "different" trials. For the easy condition, twelve S2 sequences were created with the new item in S2 differing from the item in S1 by both consonant and vowel (e.g., /lu/ instead of /mi/, Supplementary Material, Sound 3). For the difficult condition, twelve S2 sequences were created with the new item in S2 differing by the vowel only (e.g., /lu/ instead of /la/, Supplementary Material, Sound 4). When S2 was different, no substitution was made between /i/, /e/, and /y/ because of their shorter distance on the vowel triangle compared to other possible substitutions.

Speech-in-noise task

The speech-in-noise task was specifically designed for children. It was a French language adaptation (Moulin et al., 2013) of Foster & Haggard (1987)'s British Four Alternative Auditory Feature test. The adaptation for children (Bourgeois-Vionnet et al., 2020; Ginzburg et al., 2019) used pictures instead of written words and was implemented on a touch-tablet (iPad). For each trial, children had to match an aurally-presented word with its corresponding image among four pictures, by tapping on the corresponding picture on the tablet. The material consisted of 24 spoken words selected as a function of their concreteness, their frequency of occurrence in the French language and their age of acquisition (New et al., 2004). The 24 selected words were recorded several times by a French-native female speaker in order to have at least nine different sound exemplars of each word. All the exemplars were equalized in RMS amplitude and two listeners chose independently the best sounding exemplars for each word that was included in the final test. From this list of 24 words, six four-images arrays were created so that each array contained images for which the words denominating them were phonological neighbors (difficult condition, e.g. [flœʁ][bœ:ʁ][lœʁ][kœʁ]: fleur, beurre, l'heure, coeur – flower, butter, hour, heart). With the same words, six four-images arrays were created with phonologically distant words that had at least two different phonemes (easy condition, e.g. [flœʁ][mɛ̃][nø][dɑ̃]: fleur, main, nœud, dent – flower, hand, node, tooth). For each trial, an array of four-images was displayed on the touchpad's screen followed by the sound stimuli: a target word denominating one of the four images after a 600ms delay. The child had to tap with its finger on the matching image as quickly as possible but without any time limit. The next trial was triggered by the child's click on the screen. Each word was presented twice: once in the difficult condition (phonological

neighbors) and once in the easy condition (phonologically distant words) so that a total of 48 sets of four-images-array were presented. During the task, a continuous Cocktail-Party noise, made of 16 unintelligible French male and female voices, was presented binaurally via headphones. The target-word was systematically presented at 66 dB SPL. The Cocktail-party noise was presented at either 63 dB SPL (signal/noise ratio = +3 dB), or 69 dB SPL (signal/noise ratio = -3 dB).

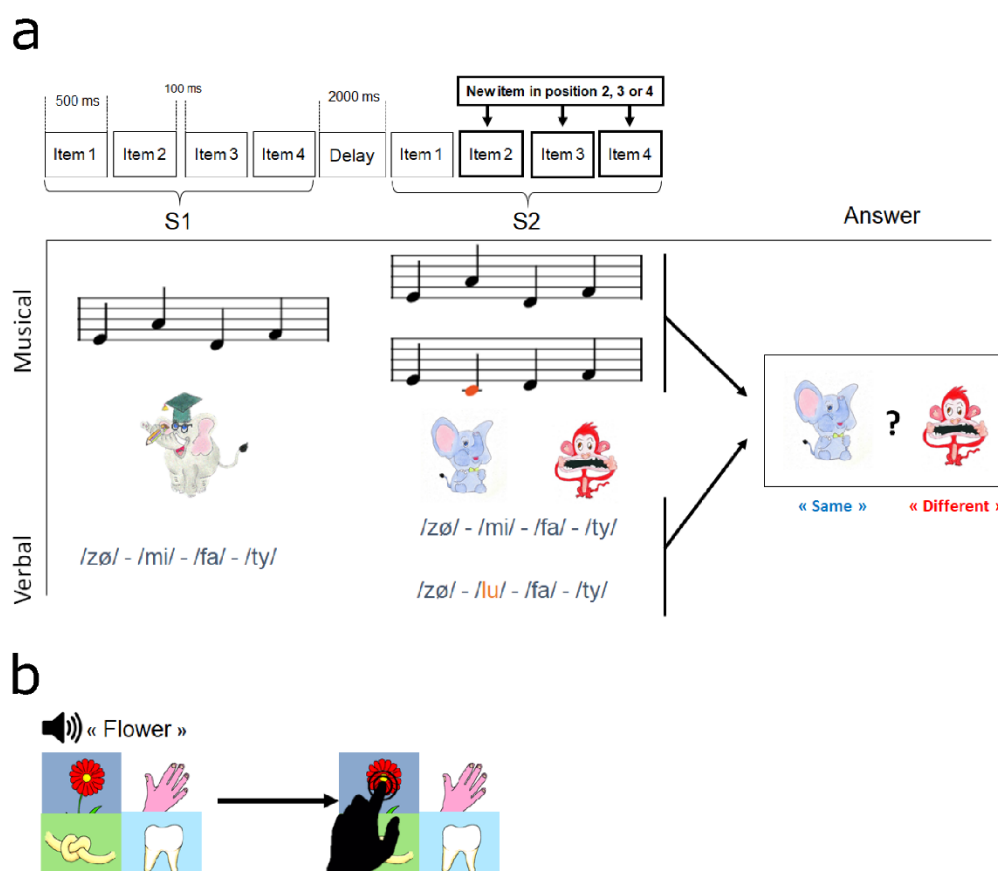


Figure 2 : (a) DMST paradigm. In each trial, the child hears the first sequence S1 and after a 2000 ms delay, the second sequence S2. In a block, in half of the trials, the S2 sequence was identical to S1, in the other half S2 was different. A new item could equiprobably appear in the second, third, or fourth position of S2. We represent here only the easy condition (Supplementary material, Sounds 1&3). In the difficult musical condition, the new item in the S2 sequence did not violate the contour of the melody (Supplementary material, Sound 2); in the difficult verbal condition, the new item in the S2 sequence changed only by the vowel (Supplementary material, Sound 4). Visual stimuli were part of the playful story children were told to understand the task, and were inspired by melodic and

rhythmic discrimination task from Ireland et al. (2018) and Wieland et al. (2015).

(b) Speech-in-noise paradigm. Children heard a word and four pictures appeared simultaneously on the touchpad. They had to tap on the corresponding picture as fast as possible. A cocktail-party noise was presented at either -30 ± 3 dB SNR in the background and two conditions of phonological proximity were possible between the words corresponding to the four images: phonologically distant (e.g., fleur-main-noeud-dent, $\backslash flœB \backslash \backslash mɛ\tilde{\backslash} \backslash nø\backslash \backslash dα\tilde{\backslash}$, as shown in the figure) or phonologically close (fleur-coeur-beurre-l'heure, $\backslash flœB \backslash \backslash kœB \backslash \backslash bœB \backslash \backslash lœB \backslash$).

Procedure

Before testing, parents filled a questionnaire about their child's level of education and their own level of education, the child's laterality, possible vision or auditory impairments, musical activities, bilingualism, learning disabilities, and eleven questions adapted from adults' musical listening questionnaires (Lévêque et al., 2018; Tillmann et al., 2014).

Children were tested by groups of five or six in the gym of their school. 3rd to 5th graders were tested first and KG to 2nd graders were tested a week after. Before testing, each child sat in front of a desk, listening to the experimenter's instructions. Before the STM tasks, the experimenter told a story about the elephant-professor and his two pupils, corresponding to the visual stimuli displayed on the computer screen during the task. These stimuli and the overall cover story for the child-friendly STM implementation were adapted from the melodic and rhythmic discrimination tasks of Ireland et al. (2018) and Wieland et al. (2015). After the six children were settled in front of their table, the experimenter sat in front of them and started telling the cover-story for the task instructions using cardboard panels on which the task visual stimuli were printed. During the first sequence, a cartoon picturing an elephant teacher would appear on the computer screen and during the second sequence, a cartoon of a nice baby-elephant would appear on the left and a cartoon of a grimacing baby-monkey would appear on the right. After the second sequence, a question mark would appear on the screen between the two cartoons. Children

were given the instruction that the baby-elephant was always repeating correctly the sequence produced by the elephant-professor and that the grimacing-monkey was always repeating incorrectly. They had to give their answer by clicking on the left button of the laptop trackpad for the baby-elephant (“same”) and on the right button for the grimacing-monkey (“different”). After the STM task, the experimenter explained the speech-in-noise task with cartoons printed on cardboard panels: children saw 4 images appear on the touchpad and at the same time, they heard a word through their headphones. They had to find the spoken word in one of the four images and tap on it as quickly as possible. They had to ignore the people talking in the background (cocktail-party noise). A training block of 4 trials was given first. The entire testing session lasted around 30 minutes, including instructions and breaks.

All children underwent:

- The Short-Term Memory (STM) recognition task with both material (verbal/musical) and difficulty levels (easy/difficult). To avoid having a testing session of more than 30 minutes, younger children (KG, 1st, and 2nd grade) were not tested with the music difficult block since they were slower than the older children and the results of older children tested during the first day revealed that this condition was the most difficult. Auditory stimuli were presented with AKG-142-HD headphones and visual stimuli via laptop computer screens.

For each material, the STM task was divided into four blocks of 24 trials (two materials and two conditions of difficulty). In each block, half of the trials were “same” trials (identical S1 and S2) and the other half were “different” trials. These 24 trials were pseudorandomly presented during a block, with the constraint that no more than three “same” or three “different” trials could appear consecutively. Children always began with the two easy blocks. Half of

the children began with the verbal material and the other half with the musical material. Each of the four STM blocks lasted around 4 minutes and was preceded by a training with corresponding material and difficulty conditions. The training consisted of 2 “same” and 2 “different” trials with a smiley-shaped error feedback for each trial. At the end of the training blocks and the test blocks, a feedback specified the number of correct answers. For the test blocks, no trial-based error feedback was given. Overall, the STM tasks lasted around 20 minutes.

- The speech-in-noise test with two levels of noise (-3 dB SNR and +3 dB SNR) and two levels of phonological proximity (Bourgeois-Vionnet et al., 2020; Ginzburg et al., 2019). Auditory stimuli were displayed through Sennheiser HD-250-Pro headphones and the task was performed on touchpads. Children underwent a training comprising eight sets of images with a signal/noise ratio (SNR) of +3 dB, then a 48-trial block (24 words in each of the two phonological conditions, in pseudo-random order, so that the same word was not presented one after the other) with a SNR of +3 dB and finally a second 48-trial block with a SNR of -3 dB. Each block lasted around 3 minutes, so the speech-in-noise test lasted around 10 minutes overall with the instructions.

Adult participants were tested in the lab with the four STM tests, as for the older children.

STM data preprocessing

To assess whether instructions were well understood, we computed the STM task’s answers at inappropriate time points, the correct response would be any answer given after S2. Responses during the S2 sequence were coded as “anticipation”, responses between the child’s response and the next trial were coded as “responses between trials” and responses during S1 or the retention

delay between S1 and S2 as “within-trial error”. Distribution of responses between trials and anticipations showed no particular information about a possible indicator of misunderstanding and/or problematic impulsivity of children during the task. However, within-trial errors displayed interesting information about children’s ability to understand the task. A high rate of errors during the S1 sequence corresponded to the experimenters’ observation of children’s inability to understand the task correctly during the experimental session. On the account that they displayed more than 5 within-trial errors out of 24 trials during at least one block, we excluded 8 children among the 100 from the analysis (two of them were in kindergarten, three were in 1st grade, one in 2nd grade, one in 4th grade and one in 5th grade).

Data analyses

In all analyses, the developmental aspects of children’s performance³ were tested as a function of children’s school grades, allowing for homogeneous groups with children being equally scholarly educated regarding reading and other abilities.

For the STM task, measures of d' and bias (c parameter) were obtained according to Signal Detection Theory (SDT) for each material (verbal/musical), for each difficulty level (easy/difficult) and for each participant (Macmillan & Creelman, 2004). Hit corresponded to a correct answer for different trials. False alarm corresponded to an incorrect answer for same trials. d' and criterion (c) were calculated using Dominique Makowski’s “dprime” R function from the *psycho* package (Makowski, 2018). d' , or sensitivity, was calculated as the z-score of False Alarms subtracted from the z-score of Hits. The criterion, c , is calculated as the mean z-score of Hits and False-alarm rates multiplied by

³ it should be noted that from this point on, the term "performance" refers to the sensitivity measure (d') described in this section

minus one and reflects an observer’s bias to say yes (in our case “different”) or no (“same”), an unbiased observer having a value around 0. A liberal bias (tendency to say “different”) results in a negative c , a conservative one results in positive c . Correction of extreme values are made following the recommendation of Hautus (1995). Furthermore, we analyzed the response times of participants after the end of S2, on correct trials. Recency effects were examined by analyzing the percentage of correct responses as a function of the position of the new item in ‘different’ trials (2nd, 3rd, or 4th position).

For the speech-in-noise task, we obtained percentages of correct answers for each child, for each condition of phonological proximity (close/distant) and for each signal/noise ratio (-3 dB and +3 dB).

Condition	KG	$1^{st} G$	$2^{nd} G$	$3^{rd} G$	$4^{th} G$	$5^{th} G$	<i>Adults</i>
Music Easy	0.26 (0.62)	0.22 (0.30)	0.34 * (0.33)	0.36* (0.29)	0.19* (0.19)	0.30* (0.26)	0.16 (0.19)
Music Difficult	NA	NA	NA	0.69* (0.28)	0.70* (0.28)	0.55* (0.36)	0.54* (0.34)
Verbal Easy	0.079 (0.63)	0.039 (0.14)	0.02 (0.25)	0.023 (0.28)	0.17 (0.21)	0.04 (0.21)	0.14* (0.13)
Verbal Difficult	0.25 (0.62)	-0.11 (0.29)	0.0074 (0.27)	0.06 (0.26)	0.18* (0.20)	0.099 (0.20)	0.14* (0.13)

Table 2 : Mean c values and standard deviation as a function of grade (columns) and condition (rows). Note: As the mean value of the c parameter of an unbiased observer is 0, we performed one-sample t -tests (Bonferroni corrected per condition) comparing c mean values per grade and condition to 0. All mean c values (except first grades in the verbal difficult condition) were numerically higher than 0 (conservative bias) as is usually observed in DMST tasks. Asterisks indicate a c value significantly different from 0.

Statistical analyses were performed on R (3.5.5 version). Mixed-design ANOVAs were performed to analyze data from the STM tasks using the *rstatix* R package (Kassambara, 2020). Due to the unbalanced experimental design (younger children did not undergo the difficult musical STM task), three sets

of analyses were performed to assess all aspects of the data (factors for each analysis are detailed in the results section). Greenhouse-Geisser correction was applied when the sphericity assumption was violated, assessed by Mauchly's test. Significant effects and interactions were analyzed using pairwise t-tests as post-hoc tests adjusted with the Tukey method. We also performed one-sampled t-tests comparing c values to 0, the value of an unbiased observer. Spearman correlations were performed between age and performance in the DMST and the speech-in-noise task. Spearman partial correlation analyses with age as control variable were performed between performance in the STM task and the speech-in-noise task using the *ppcor* R package (Kim, 2015).

Two ANOVAs are presented in the results section: the first one concerns children of all six grades (KG to 5th grade) on the easy conditions of the STM task in order to examine the developmental trajectory of STM for both materials. The second one concerns older children (from 3rd grade to 5th grade) and examines the differential effects of material and its interaction with the task difficulty. A third ANOVA has been performed with children of all six grades on the verbal conditions in order to assess difficulty effects and is provided as a supplementary figure (see supplementary Figure S2 and Table S1).

Results

Musical and verbal STM: easy conditions

We computed two 6x2 mixed-design analyses of variance (ANOVA) for the easy conditions on d' (Figure 3) and c (Table 2) with Grade as a between-subjects factor (kindergarten to 5th grade) and with Material as within-subject factor (musical/verbal), as well as a 6x2x2 mixed-design ANOVA for response times analysis (Figure 3) with the aforementioned factors and with type of trial as within-factor (same/different). We also performed a 6x2x3 ANOVA on the

percentage of correct responses in different trials to assess recency effects, with Grade as between-subject factor, and Material and Position (2nd, 3rd, and 4th position) as within-subject factors. The position factor corresponded to the position of the item that changed during the S2 of different trials. Complete ANOVA results are shown in Tables

Table 3, and we present significant effects and interactions below.

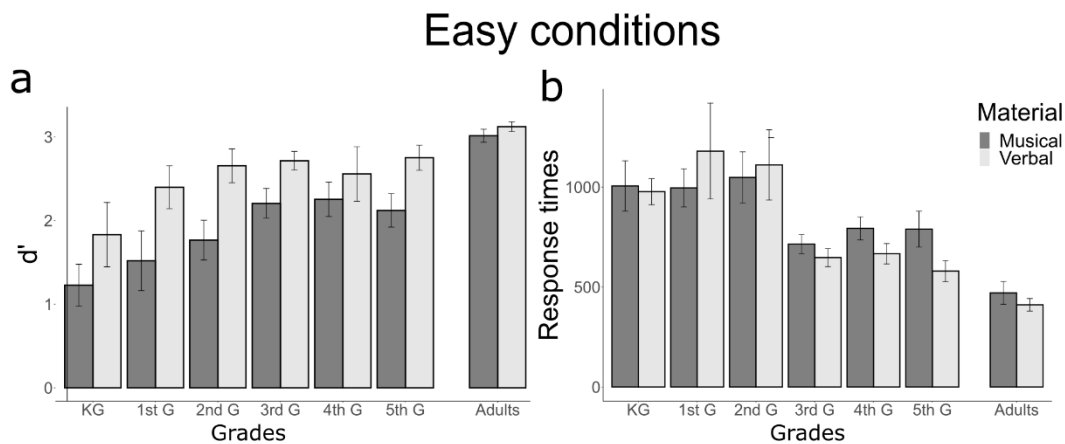


Figure 3 : Performance for the easy conditions as a function of material (musical or verbal) for all children’s grades (KG, kindergarten; G, grade) and for adults. (a) Mean and standard error of children and adult’s sensitivity (calculated as the d'). (b) Mean and standard error of children and adult’s correct response times (time in millisecond that subjects spent after the end of S2 before giving a “same” or “different” answer)

d' . The main effect of Grade was significant, $F(5,86) = 3.030$, $p = .014$, $\eta_p^2 = .257$, revealing better performance for older children. Post-hoc tests showed a significantly lower performance of kindergarten children compared to all other grades (all p -values $< .02$) except 1st grade and 2nd grade (both p -values $> .13$). The main effect of Material was significant $F(1,86) = 32.33$, $p < .001$, $\eta_p^2 = .273$, showing better performance with the verbal material than with musical material. The interaction between grade and material was not significant ($p = .644$) even though the observation of the results suggests that 3rd graders performed better for the musical material than 2nd graders whereas

for the verbal material 2nd graders seemed to reach a similar performance level as did 3rd graders (Figure 3).

Correct response times. There were no significant effects in the response time analysis. We expected a grade effect which was marginally significant, $F(5,85) = 2.238$, $p = .058$, $\eta_p^2 = .116$.

Bias. Comparison of the mean c parameter per grade and condition to zero showed that 2nd, 3rd, 4th and 5th grade children displayed a positive response bias c that was significantly above zero in the easy musical condition (Table 2). The ANOVA revealed a significant effect of Material $F(1,86) = 40.89$, $p < .001$, $\eta_p^2 = .322$, with a higher c for the musical material than the verbal material.

Recency effects. Only the main effect of Material was significant, $F(1,87) = 54.850$, $p < .001$, $\eta_p^2 = .387$, with better performance for verbal than musical material, thus mirroring the d' analysis.

When children were gathered in three groups of age (KG-1st, 2nd-3rd, 4th-5th), the ANOVA on recency effect showed a main effect of Grade, albeit with a small effect size $F(2,89) = 3.591$, $p = .032$, $\eta_p^2 = .075$.

Musical and verbal STM: effect of difficulty and material

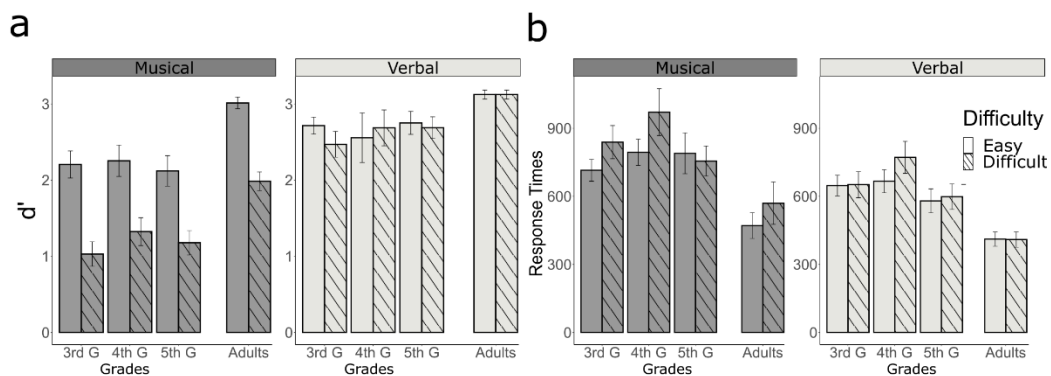


Figure 4 : Performance for all conditions of Material (musical or verbal) and Difficulty (easy or difficult) for children of third, fourth, and fifth grades and for adults. (a) Mean and standard error of children and adult's sensitivity (calculated as the d'). (b) Mean and standard error of children and adult's response time (time in millisecond that subjects spent before giving a "same" or "different" answer)

Two 3x2x2 mixed-design analyses of variance (ANOVA) were performed on d' (Figure 4) and c (Table 4) with Grade as a between-subjects factor (3rd to 5th grade), and Material (verbal/musical) and Difficulty (easy/difficult) as within-subject factors. A 3x2x2x2 mixed-design ANOVA was performed on response times (Figure 4) with the aforementioned factors and with type of trial as within-subject factor (same/different). A 3x2x2x3 mixed design ANOVA was performed on the percentage of correct response for different trials, adding the Position factor (2nd, 3rd and 4th position) as within-subject factor. These analyses were only possible with data from children between 3rd and 5th grade ($n = 50$).

d' . No significant effect of Grade was observed in these groups spanning a more restricted age range, $F(2,47) = 0.325$, $p = .7$, $\eta_p^2 = .038$. We observed a main effect of Material, $F(1,47) = 94.8$, $p < .001$, $\eta_p^2 = .787$, with lower performance for musical material than for verbal material. The main effect of Difficulty was significant, $F(1,47) = 31.24$, $p < .001$, $\eta_p^2 = .544$ and interacted with Material, $F(1,47) = 42.4$, $p < .001$, $\eta_p^2 = .474$: performance was lower in the difficult condition than in the easy condition for the musical material ($p < .001$), but not for the verbal material ($p = .96$).

Correct response times. We found a significant effect of Material $F(1,47) = 22.373$, $p < .001$, $\eta_p^2 = .428$ with longer response times for the musical material than for the verbal material. The material effect interacted significantly with the type of trial factor $F(1,47) = 7.874$, $p = .007$, $\eta_p^2 = .143$ with a difference between musical and verbal materials only for different trials ($p < .001$).

Bias. Comparison of the mean c parameter per grade and condition to zero showed that all children from 3rd to 5th grade displayed positive response bias c that was significantly above 0 in the difficult musical condition (Table 2). The ANOVA revealed a significant effect of material $F(1,47) = 92.585$, $p <$

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

.001, $\eta_p^2 = .734$ with a higher positive c for the musical material than the verbal material. The main effect of difficulty was significant, $F(1,47) = 40.558$, $p < .001$, $\eta_p^2 = .432$, children displaying a higher c for the difficult condition. The material and difficulty interaction was significant $F(1,47) = 23.118$, $p < .001$, $\eta_p^2 = .330$. Post-hoc tests revealed a significantly higher c in the difficult condition compared to the easy condition, only for the musical material ($p < .001$) but not for the verbal material ($p = 0.46$).

Recency effects. We found a Position effect $F(2,144) = 32.381$ $p < .001$, $\eta_p^2 = .403$ with performance on the 2nd position lower than for 3rd position ($p = .035$) and 4th position ($p < .001$) and 3rd position lower than 4th ($p = .028$). The Position, Material, and Difficulty interaction $F(2,144) = 38.196$, $p < .001$, $\eta_p^2 = .442$ showed a lower performance on the 2nd position compared to the 3rd and 4th and between 3rd and 4th only for the difficult musical material ($p < .001$).

Partial correlations between STM and speech-in-noise performance

For the speech-in-noise task, as expected, effects of age, phonological proximity, and noise intensity were observed (Ginzburg et al., 2019). The correlation analysis shown here-after include 87 children: data from five children were excluded because of technical difficulties when recording speech-in-noise data.

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

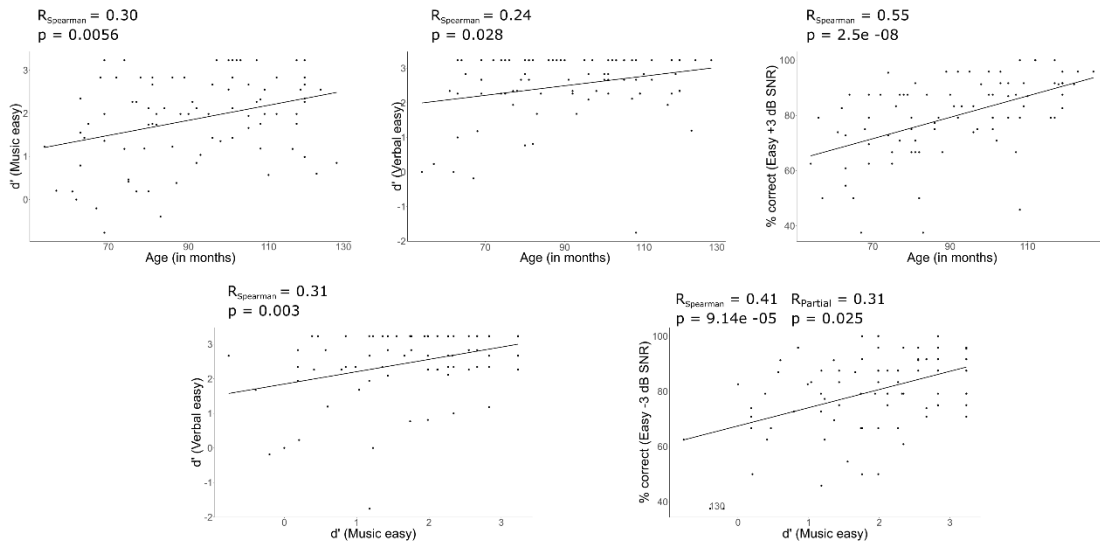


Figure 5 : Top panels: Spearman correlation between children’s age (in number of months) and the performance of children for the STM task (calculated as the d') in the musical easy condition (top left), the verbal easy condition (top middle) and the performance of children in the speech-in-noise task (in percentage of correct response) for the easy condition and a -3 dB SNR (top right). Lower left panel: correlation between the musical easy condition and the verbal easy condition in the STM task. Lower right panel: partial correlation between the musical easy condition in the STM task and the speech-in-noise task in the easy phonological condition and -3 dB SNR. $R_{Spearman}$: Spearman correlation coefficient and its corresponding p -value. $R_{partial}$: partial correlation coefficient after factoring out the effect of age and its corresponding p -value. Partial correlations were Bonferroni corrected for eight comparisons (two STM conditions and four speech-in-noise conditions): the p -values were multiplied by eight. The lower right panel correlation was the only correlation between STM performance and Speech-in-Noise performance that showed significance when a partial correlation was computed with age as a control variable

We first performed correlations between musical and verbal STM performance (d' , in the easy condition only, as data for the difficult condition was not available for children from KG to 2nd grade) and the age of children (in months). Both conditions were significantly correlated with age $r(85) = .30$, $p = .0056$ for the musical material and $r(85) = .24$, $p = .028$ for the verbal material and these two correlation coefficient were not significantly different ($p = .80$). We then performed partial correlation accounting for the age variable between musical and verbal STM performance (d'), in the easy condition only,

as data for the difficult condition was not available for children from KG to 2nd grade. Musical and verbal STM were significantly correlated even after factoring out the effect of age, $r(85) = .26$, $p = .014$. Next, we performed partial correlations between performance in the STM tasks (easy conditions) and in the speech-in-noise tasks accounting for the age variable. The analysis revealed that only the correlation between music STM performance with performance in the phonologically distant condition with a -3 dB SNR in the speech-in-noise task was significant, $r(85) = 0.32$, $p = 0.025$, after Bonferroni correction for multiple comparisons (Figure 6). No other significant correlations were observed between performance in STM tasks and speech-in-noise tasks when age was considered ($p > 0.53$).

Effect of learning disorders or musical training

Five children, out of the 92 included in the analyses above, had various learning disorders diagnosis (dyslexia, dysphasia, dyscalculia, dysorthographia, dysgraphia), according to parental reports. No statistical analysis was conducted on their performance because of their small number. Nonetheless, for the STM task, we compared each child's performance in every condition to the corresponding median and their placement in the quartiles of performance of other children of the same grade and in the same condition. This analysis revealed that in the musical easy condition, all five children with learning disorders scored below the median. Among those five children, four of them performed in the first quartile. In the other conditions, at most two out of these five children performed above the median. All the ANOVAs described before were performed excluding these five children and, aside from a small decrease of statistical power, the effects remained consistent.

For the speech-in-noise task, we compared these children's performance in each condition of phonological difficulty (easy/hard) and noise intensity

(SNR-3/SNR+3) to the median of the other children for the corresponding grade, phonological difficulty, and noise intensity. Noticeably in the easy condition (phonologically distant) with a +3 dB SNR, all of the five children with learning disorders performed below the median, in the lowest quartile. In both difficulty conditions for the -3 dB SNR, at most one child performed above the median, but never the same child. In the difficult condition for +3 dB SNR, only two children performed below the median.

Furthermore, nine children reported musical training of at least one year (conservatory or communal music school). Given their small number, we did not perform any statistical analysis but when we removed them from the analysis, the effects remained consistent. Seven out of these nine children performed above the median and five out of these seven performed in the upper quartile in the music easy condition and in the verbal difficult condition. Five out of these nine children underwent the musical difficult condition (the others being too young) and all of them performed above the median for this condition. For the speech-in-noise task, a majority of children with musical training (at least seven out of nine in each condition) scored above the median in all conditions.

Comparison between children and adults

To compare performance of the oldest children with performance of adults, three 2x2x2 mixed-design ANOVA on d' , RTs and c were performed with Group (adults ($n = 12$) and children (5th graders, $n = 17$)) as between-subject factor, and Material (verbal/musical) and Difficulty (easy/difficult) as within-subject factors and type of trials as within-subject factor (same/different) for the response times analysis. A 2x2x2x3 mixed-design ANOVA was performed on the percentage of correct responses for different

trials adding the Position factor (2nd, 3rd and 4th position) as within-subject factor. Complete results are shown in Table 5.

d'. The main effect of Group was significant, $F(1,27) = 13.25$ $p = .00114$, $\eta_p^2 = .743$ with children having poorer performance than adults. A significant Group and Material interaction was found $F(1,27) = 4.981$, $p = .0341$, $\eta_p^2 = .269$. Post-hoc tests showed a significant difference between all comparisons ($p < .003$) except for a marginally significant difference between children and adults for the verbal material ($p = .061$). Adults displayed a significantly smaller difference between musical and verbal material than children ($p = .02$, t-test comparing the subtraction of musical and verbal performance between adults and children). All other effects mirrored the former *d'* analysis on children's data.

Correct response times The main effect of Group was significant $F(1,27) = 4.825$, $p = .037$ $\eta_p^2 = .152$ with longer response times for children than for adults. The main effect of Material was significant $F(1,27) = 9.911$, $p = .004$ $\eta_p^2 = .269$ with longer response times for the musical material. There was also a significant interaction between Material and Type of trial $F(1,27) = 8.625$, $p = .007$, $\eta_p^2 = .242$ with a difference between musical and verbal material only in different trials ($p < .001$).

Bias. Comparison of the mean *c* parameter per condition to 0 showed that adults displayed a positive bias parameter *c* for all conditions and all of them except the musical easy condition were significantly different from zero (Table 2). No effect of group was found in the ANOVA.

Recency effects. We observed an effect of the Group factor $F(1,84) = 10.99$, $p = .00254$, $\eta_p^2 = .177$ with lower performance for children. None of the other significant effects or interactions involved the Group or the Position factor (Figure 5).

We emphasize the fact that ceiling performance (e.g. 100% correct responses) was observed for the majority of adults in the easy and difficult conditions for the verbal material and for the easy musical condition. However less than half of 5th graders reached ceiling in all conditions.

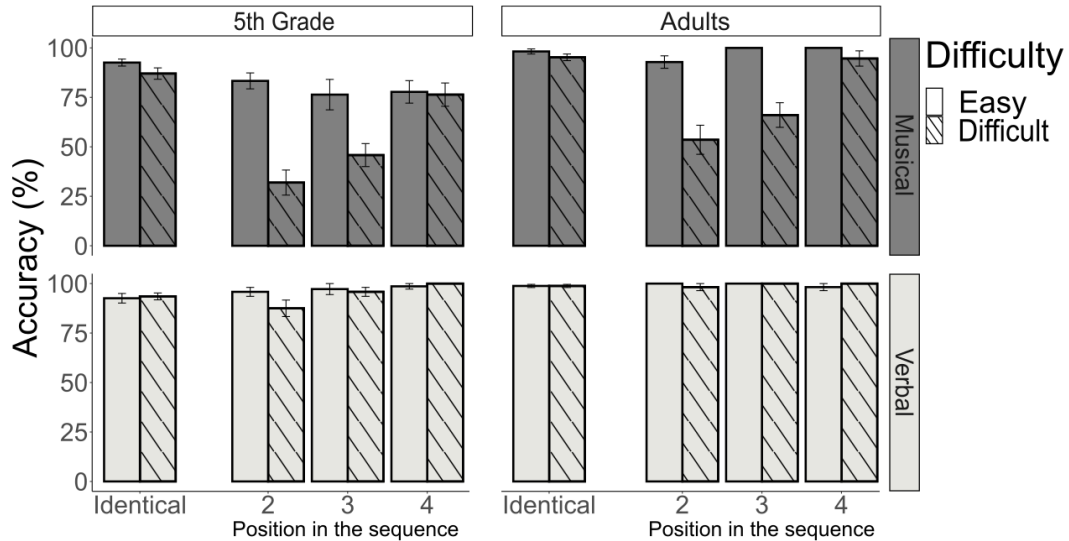


Figure 6 : Recency effects for the older children (fifth grade, $n = 17$) and adults (% of correct responses for different trials). Results are presented as a function of the material musical/verbal), difficulty (easy/difficult) and the position of the item change in the S2 sequence (second, third, or fourth position). Percentage of correct response for identical S1-S2 are also represented for comparison. These effects were similar for the other younger children

Discussion

The aim of this study was to investigate the development of auditory STM in 5 to 10-year-old children for musical and verbal material. A DMST was created in a child-friendly version in highly comparable ways for both materials, each one with two levels of difficulty. Results showed that overall, younger children (KG and 1st grade, see Table 1 for age equivalences) displayed poorer memory performance than older children (2nd, 3rd, 4th and 5th grade). This was observed for both materials: children's performance increased from KG to 2nd grade, and then remained stable until 5th grade. Overall, children's performance was lower for musical material (sequences of 4 tones) than verbal material

(sequences of 4 syllables). Adults showed the same pattern of performance, but with ceiling performance for verbal material. For both musical and verbal materials, children did not reach the same level of performance as adults, thus revealing the immaturity of auditory STM even in 10-year old children. Response times analysis showed the same pattern as performance analysis: younger children were slower than older children thus confirming the absence of an impulsivity issue or a potential difference in speed-accuracy trade-off between age-groups. An effect of difficulty level was found for the musical material, revealing a facilitatory effect of contour violation and large interval changes, in both children and adults. Manipulating the difficulty in the verbal material (i.e., changing only a vowel in the second sequence in the difficult task instead of both consonant and vowel in the easy version) elicited only minor decrease in performance, if any. As for the recency effects expected in that kind of task, they were more apparent in the difficult conditions. Children displayed a recency effect for both materials, but older children and adults mostly for the difficult musical material.

Developmental trajectory of auditory STM

The present study, based on a recognition paradigm with a fixed number of items, suggests the following pattern of development for musical and verbal STM: increasing performance from KG to 2nd grade (5 to 7 years-old) followed by a standstill until 5th Grade (10 years old). However, neither Gathercole et al. (2004b) nor Alloway et al. (2006) observed such a pattern for verbal STM (for a comparison between the former two studies and the present one, see Figure 7), they rather observed a linear increase in capacity. Regarding musical STM, Clark et al. (2018) found a developmental increase in children from 6 to 13 years-old similar to the one of visuo-spatial STM, with a linear increase in

storage capacity as observed in former studies (Alloway et al., 2006; Gathercole et al., 2004).

The difference of developmental pattern observed between our study and the former ones may rely on the difference of cognitive processes used in recall tasks and DMST. Recall tasks might rely more strongly on memory capacity and long-term knowledge (if it is used with digits or words), whereas the DMST task might involve different procedural processes as it involves a comparison process between S1 and S2 and rely more on sensory memory traces as they can be used in stimulus recognition (Cowan, 1984). Regarding the weight of long-term knowledge, lexical knowledge has indeed been found to have a significant influence on verbal STM in children when recall tasks are used (Gathercole et al., 2001; Messer et al., 2015) and also when serial-order STM reconstruction tasks are used (Gorin et al., 2018; Leclercq & Majerus, 2010; Majerus et al., 2006). The reconstruction task is supposed to minimize phonological and lexical demands, however these studies are using verbal materials (like animal names in Leclercq & Majerus, 2010 for example). Forward recall tasks also rely on phonological production and thus present constraints linked to the phonological loop and specific mechanisms of motor production.

Several factors have been proposed to account for the increase of performance in studies investigating the development of STM: the decrease of memory trace decay throughout childhood, the increase of memory capacity, the increasing involvement of executive functions as they mature, and the influence of long-term knowledge that would support short-term storage (Messer et al., 2015). The developmental trajectory observed in the present study could thus arise from the use of a DMST that reduces lexical demands and necessitates different executive demands during the comparison between

the two sequences. Furthermore, DMST has the significant advantage of allowing the same implementation for different auditory material (verbal, musical, timbre), which allows direct comparisons.

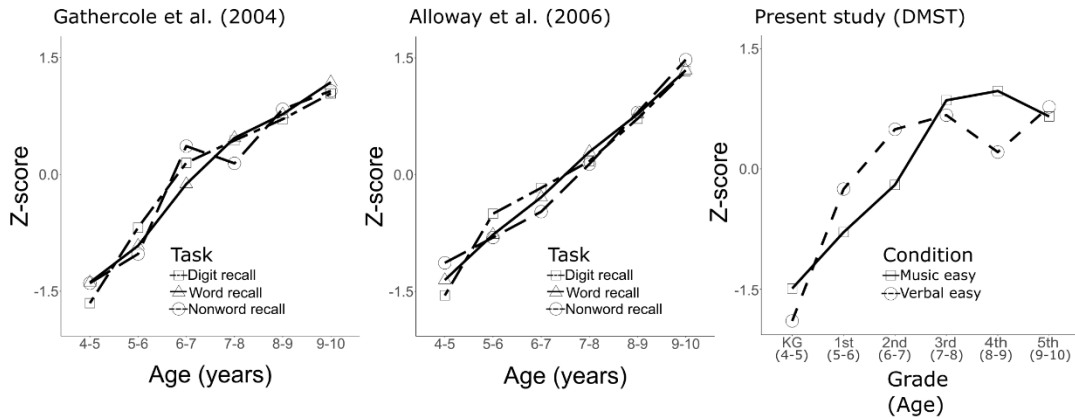


Figure 7: Left and middle panels: z-scores of mean performance in verbal forward recall tasks with digits, words, and nonwords in children from 4 to 10 years of age, adapted from Gathercole et al. (2004) and Alloway et al. (2006), respectively. Right panel: z-scores of mean musical and verbal recognition tasks of the present study in the easy condition

Development of musical and verbal STM: shared and distinct mechanisms

As mentioned in the introduction, several WM model and in particular their account for short-term storage could be divided into two groups regarding their prediction for the developmental trajectory of musical and verbal STM. On one hand, theoretical framework based on Baddeley and Hitch (1974) WM model postulate the existence of a separate short-term storage for musical material (Pechmann & Mohr, 1992) and even a specific musical central executive (Ockelford, 2007). These frameworks might predict two different developmental trajectories for musical and verbal STM as they would rely on different attentional processes. On the other hand, other models consider that attentional processes involved in the maintenance of items in STM do not present such a modular structure (Barrouillet & Camos, 2007; Cowan et al., 1998). A common consideration between these models is that maintenance processes involving attentional resources used to refresh memory traces in short-

term storage are non-modular and can be allocated to different materials. Such models account for the distinction between domain-general and domain-specific processes in musical and verbal STM (Gorin et al., 2018) and they would probably account for a similar developmental trajectory between musical and verbal STM as it was observed in this study. In addition to a similar trajectory, we found recency effects for both materials. Some models posit that recency effect would represent behavioral signatures of serial order constructs in recall and recognition tasks (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). Indeed, it has been argued that musical and verbal STM systems may rely on similar sequential processes. These similar sequential processes were characterized in adults by Gorin et al (2018) who found similar selective sensitivity to time-based interference in musical and verbal STM (with a mixed recall/recognition paradigm) and similar transposition gradients. They also found similar error patterns, sequence length effects, and recency and primacy effects for both materials as well as similar limited capacity and an effect of pitch proximity, comparable to phonological proximity (Williamson et al., 2010). Another finding in favor of shared mechanisms between the two materials is the significant correlation between musical and verbal STM performance, even when the effect of age was factored out.

We observed similarities in the developmental trajectory for both materials, but we also observed better performance for the verbal material, despite using the same number of items for both tasks, both for children and adults (Figure 4). Although this latter finding could be related to discriminability differences for musical and verbal items used in the DMST, we cannot rule out the possibility that modality-specific STM systems could treat the two materials differently, as domain-specific systems have been suggested for encoding, storage, and maintenance of musical and verbal information

(Gorin et al., 2018; Schulze & Tillmann, 2013). In addition, the fact that only musical STM performance correlated with the speech-in-noise performance in the present study suggests differentiated encoding processes between musical and verbal STM. A study involving comparison between STM for words, tones, and timbres using a recognition paradigm in adults, suggested similar storage of musical and verbal information and different internal sensorimotor codes used to maintain musical and verbal information (Schulze & Tillmann, 2013). In our present study, the similarity of the developmental trajectory for both materials could be the consequence of the involvement of shared domain-general systems in serial order coding, given in particular the finding of a recency effect for both materials, shared maintenance processes, and similar rehearsal mechanisms. Conversely, the difference of performance that we found in children between materials, might arise from different sensorimotor codes used for musical and verbal information.

Development of auditory cognition

The secondary aim of this study was to investigate auditory cognition beyond auditory STM. Indeed, the links between CAPDs and learning disorder, as well as the advantage of musicians in speech-in-noise abilities (Parbery-Clark et al., 2009; Slater et al., 2015; Zendel et al., 2015), suggest that the joint investigation of speech-in-noise and verbal and non-verbal STM would be informative. Here, children underwent a speech-in-noise task (Ginzburg et al., 2019) and it was found that only performance in the musical DMST correlated with speech-in-noise performance, specifically in the phonologically easy condition with a -3 dB SNR (Figure 5). This finding suggests that children might rely on perceptual processes used in pitch encoding to process speech in cocktail-party noise, these processes being shared with musical STM but not with verbal STM. These results verify the already observed link between sound segregation

and pitch information processing (Oxenham, 2008). Overall, these results seem to corroborate the hypothesis that musical and verbal STM have distinct mechanisms regarding item-information processing. Indeed, we showed that domain-general procedural and attentional processing seem to be involved in both musical and verbal STM given their similar developmental trajectory. The distinct mechanisms involved in musical and verbal STM appears lie in item-specific encoding processes that only speech-in-noise and musical STM share in the present study.

Auditory cognition and learning disorders

These two components of auditory cognition (STM and speech-in-noise) might thus share processes that are of particular interest to understand the underpinnings of central auditory processing. We found, in the present study, that children who presented learning disorders all performed below the median in the musical easy condition of the DMST and in the easy condition at +3 dB SNR of the speech-in-noise task. These results indicate that the musical easy condition in the STM task and the easy +3 dB SNR condition in the speech-in-noise task might allow discriminating children with learning disorders. These results are in agreement with the hypothesis of a general deficit in auditory perception in learning disorders (Nithart et al., 2009) encompassing verbal STM deficits (Perez et al., 2012) and speech-in-noise deficits (Ziegler et al., 2005). Therefore, it seems that STM for pitch and speech in cocktail party noise could be of great interest to identify early CAPD and, as suggested by these studies, early learning disorders. As comorbidity between CAPD and learning disorders has been reported, the use of early identification of CAPD would facilitate the adaptation of a child's school and home environment earlier in development. Indeed, the diagnosis of learning disorders is currently highly reliant on reading abilities and thus cannot be done before reading acquisition. The use of auditory

child-adapted tasks, such as the ones used here, allows overriding the reading acquisition problematic as no reading abilities are required. Furthermore, remediation and clinical treatment could benefit from these insights as pitch encoding seems to be closely related to learning disorders. Indeed, clinical population could take advantage of the indirect effect of pitch encoding enhancement that arise from musical training (Forgeard et al., 2008).

Conclusion

This study investigated the developmental trajectory of auditory STM in 5 to 10 years-old children for musical and verbal material. Results suggest that auditory STM is developing throughout childhood and is still under maturation at the age of 10. With the use of a recognition task, we observed that STM performance increases until 2nd grade and levels-off until 5th grade for musical and verbal material. Using the DMST allowed discussing shared processes between the two materials and the different processes at stake in recognition tasks compared to forward recall tasks. Children, as well as adults, also showed poorer performance for the musical material compared to the verbal material, providing some evidence for specific mechanisms for the processing of the two materials in STM. We also observed that similar processes might be at stake between musical STM and speech-in-noise perception revealing the relevance of investigating the specific processes involved in musical information processing. Future studies should further investigate the development of auditory cognition in a systematic way that relies less on verbal production and LTM knowledge as serial recall tasks do. Future investigation should also assess jointly short-term memory and speech-in-noise processing in children with typical development and learning disorders, to pave the way for new diagnosis and rehabilitation tools of central auditory processing deficits.

References

- Adlard, A., & Hazan, V. (1998). Speech Perception in Children with Specific Reading Difficulties (Dyslexia). *The Quarterly Journal of Experimental Psychology Section A*, *51*(1), 153-177. <https://doi.org/10.1080/713755750>
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children : Are They Separable? *Child Development*, *77*(6), 1698-1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- Atterbury, B. W. (1985). Musical Differences in Learning-Disabled and Normal-Achieving Readers, Aged Seven, Eight and Nine. *Psychology of Music*, *13*(2), 114-123. <https://doi.org/10.1177/0305735685132005>
- Attout, L., & Majerus, S. (2015). Working memory deficits in developmental dyscalculia : The importance of serial order. *Child Neuropsychology*, *21*(4), 432-450. <https://doi.org/10.1080/09297049.2014.922170>
- Baddeley, A., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation*, *8*, 47-90.
- Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Éds.), *The Cognitive Neuroscience of Working Memory* (p. 59-80). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198570394.003.0004>
- Berz, W. L. (1995). Working Memory in Music : A Theoretical Model. *Music Perception: An Interdisciplinary Journal*, *12*(3), 353-364. <https://doi.org/10.2307/40286188>
- Bourgeois-Vionnet, J., Moulin, A., Hermier, M., Pralus, A., Tillmann, B., Caclin, A., & Nighoghossian, N. (2020). A case of verbal and emotional prosody processing dissociation after a right temporal venous infarct. *Neurological Sciences*, *41*(6), 1615-1618. <https://doi.org/10.1007/s10072-020-04259-y>
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). *Speaking Clearly for Children With Learning Disabilities Sentence Perception in Noise*. 18.
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory : Insights from neurodevelopmental and neurological disorders: Musical and verbal short-term memory. *Annals of the New York Academy of Sciences*, *1423*(1), 155-165. <https://doi.org/10.1111/nyas.13733>
- Chandrasekaran, B., & Kraus, N. (2010). Music, Noise-Exclusion, and Learning. *Music Perception*, *27*(4), 297-306. <https://doi.org/10.1525/mp.2010.27.4.297>
- Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and Spatial Short-Term Memory : Common Sources of Developmental Change? *Journal of Experimental Child Psychology*, *73*(1), 7-44. <https://doi.org/10.1006/jecp.1999.2493>
- Clark, K. M., Hardman, K. O., Schachtman, T. R., Sauls, J. S., Glass, B. A., & Cowan, N. (2018). Tone series and the nature of working memory capacity development. *Developmental Psychology*, *54*(4), 663-676. <https://doi.org/10.1037/dev0000466>

- Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia*, *155*, 107811. <https://doi.org/10.1016/j.neuropsychologia.2021.107811>
- Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, *36*(1-2), 1-17. <https://doi.org/10.1080/02643294.2019.1578205>
- Cowan, N. (1984). On Short and Long Auditory Stores. *Psychological Bulletin*, *96*(2), 341-370. <https://doi.org/10.1037/0033-2909.96.2.341>
- Cowan, N. (1998). *Attention and Memory: An Integrated Framework*.
- Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research* (Vol. 169, p. 323-338). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The Role of Attention in the Development of Short-Term Memory: Age Differences in the Verbal Span of Apprehension. *Child Development*, *70*(5), 1082-1097.
- Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). *Two Separate Verbal Processing Rates Contributing to Short-Term Memory Span*. 20.
- Dempster, F. N. (1981). Memory Span: Sources of Individual and Developmental Differences. *Psychological Bulletin*, *89*(1), 63-100.
- Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & Psychophysics*, *35*(5), 477-488. <https://doi.org/10.3758/BF03203924>
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 383-390. <https://doi.org/10.1525/mp.2008.25.4.383>
- Foster, J., & Haggard, M. (1987). The Four Alternative Auditory Feature test (FAAF)—Linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology*, *21*(3), 165-174. <https://doi.org/10.3109/03005368709076402>
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410-419. [https://doi.org/10.1016/S1364-6613\(99\)01388-1](https://doi.org/10.1016/S1364-6613(99)01388-1)
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177-190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable Lexical and Phonological Influences on Serial Recognition and Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 1-30. <https://doi.org/10.1080/02724980042000002>

- Ginzburg, J., Fornoni, L., Popieul, J., Pierre, C., Caclin, A., & Moulin, A. (2019). Neurodevelopmental aspects of hearing-in-noise in 5 to 11 year-old children. Oral communication at the 14th congress of the European Federation of Audiological Societies, 22-25 may 2019, Lisbon (Portugal). *Journal of Hearing Science*, *9(1)*(1), Supplement 41.
- Gordon, E. E. (1986). Manual for the Primary Measures of Music Audiation and the Intermediate Measures of Music Audiation. *Chicago: G.I.A Publications, Inc.*
- Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-Generalitv of Timing-Based Serial Order Processes in Short-Term Memory : New Insights from Musical and Verbal Domains. *PLOS ONE*, *11*(12), e0168699. <https://doi.org/10.1371/journal.pone.0168699>
- Gorin, S., Mengal, P., & Majerus, S. (2018). A comparison of serial order short-term memory effects across verbal and musical domains. *Memory & Cognition*, *46*(3), 464-481. <https://doi.org/10.3758/s13421-017-0778-0>
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3-10. <https://doi.org/10.1016/j.tics.2010.10.001>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46-51. <https://doi.org/10.3758/BF03203619>
- Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains : An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339-373. <https://doi.org/10.1037/a0034221>
- Iliadou, V., & Iakovides, S. (2003). Contribution of psychoacoustics and neuroaudiology in revealing correlation of mental disorders with central auditory processing disorders. *Annals of General Hospital Psychiatry*, *5*.
- Ireland, K., Parker, A., Foster, N., & Penhune, V. (2018). Rhythm and Melody Tasks for School-Aged Children With and Without Musical Training : Age-Equivalent Scores and Reliability. *Frontiers in Psychology*, *9*, 426. <https://doi.org/10.3389/fpsyg.2018.00426>
- Kassambara, A. (2020). rstatix : Pipe-Friendly Framework for Basic Statistical Tests. *R version 0.5.0*. <https://CRAN.R-project.org/package=rstatix>
- Keller, T. A., & Cowan, N. (1994). Developmental Increase in the Duration of Memory for Tone Pitch. *Developmental Psychology*, *30*(6), 855-863.
- Kim, S. (2015). ppcor : An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665-674. <https://doi.org/10.5351/CSAM.2015.22.6.665>
- Kramer, J. H., Knee, K., & Delis, D. C. (2000). Verbal Memory Impairments in Dyslexia. *Archives of Clinical Neuropsychology*, *15*(1), 83-93. <https://doi.org/10.1093/arclin/15.1.83>
- Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development : Evidence from a longitudinal study. *Developmental Psychology*, *46*(2), 417-427. <https://doi.org/10.1037/a0018540>

- Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in congenital amusia : Impaired recognition, but preserved emotional intensity. *Neuropsychology*, *32*(7), 880-894. <https://doi.org/10.1037/neu0000461>
- Lewandowsky, S., & Farrell, S. (2008). Short-Term Memory : New Data and a Model. In *Psychology of Learning and Motivation* (Vol. 49, p. 1-48). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)00001-7](https://doi.org/10.1016/S0079-7421(08)00001-7)
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia : The Importance of Serial Order. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.01522>
- Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory : The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*(2), 95-119. <https://doi.org/10.1016/j.jecp.2005.07.005>
- Makowski, D. (2018). The psycho Package : An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. <https://doi.org/10.21105/joss.00470>
- Männel, C., Meyer, L., Wilcke, A., Boltze, J., Kirsten, H., & Friederici, A. D. (2015). Working-memory endophenotype and dyslexia-associated genetic variant predict dyslexia phenotype. *Cortex*, *71*, 291-305. <https://doi.org/10.1016/j.cortex.2015.06.029>
- Medwetsky, L. (2011). Spoken Language Processing Model : Bridging Auditory and Language Processing to Guide Assessment and Intervention. *Language, Speech, and Hearing Services in School*, *42*, 286-296. [https://doi.org/10.1044/0161-1461\(2011/10-0036\)](https://doi.org/10.1044/0161-1461(2011/10-0036))
- Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., & Leseman, P. P. M. (2015). Growth of verbal short-term memory of nonwords varying in phonotactic probability : A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language*, *84*, 24-36. <https://doi.org/10.1016/j.jml.2015.05.001>
- Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, *41*(6), 576-600. <https://doi.org/10.3758/BF03210491>
- Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A. (2010). Nature of Auditory Processing Disorder in Children. *PEDIATRICS*, *126*(2), e382-e390. <https://doi.org/10.1542/peds.2009-2826>
- Moulin, A., Garcia, S., Jeanvoine, A., & Richard, C. (2013). French “Audimots” : Development of a French version of the Four Alternative Auditory Features Test. 11th European Federation of Audiology Societies (EFAS) Congress. *Otorhinolaryngologia Hungarica*, *59*(2), 98-99.
- Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance.

- Psychology and Aging*, 15(2), 323-334. <https://doi.org/10.1037/0882-7974.15.2.323>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516-524. <https://doi.org/10.3758/BF03195598>
- Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading Disabilities in SLI and Dyslexia Result From Distinct Phonological Impairments. *Developmental Neuropsychology*, 34(3), 296-311. <https://doi.org/10.1080/87565640902801841>
- Ockelford, A. (2007). A Music Module in Working Memory? Evidence from the Performance of a Prodigious Musical Savant. *Musicae Scientiae*, 11(2_suppl), 5-36. <https://doi.org/10.1177/10298649070110S202>
- Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*, 8(6), 537-548.
- Oxenham, A. J. (2008). Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants. *Trends in Amplification*, 12(4), 316-331. <https://doi.org/10.1177/1084713808325881>
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear and Hearing*, 30(6), 653-661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, 20(3), 314-320. <https://doi.org/10.3758/BF03199668>
- Peretz, I., & Babai, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, 30(3), 277-292. [https://doi.org/10.1016/0028-3932\(92\)90005-7](https://doi.org/10.1016/0028-3932(92)90005-7)
- Perez, T. M., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a Specific Impairment of Serial Order Short-term Memory in Dyslexic Children: Serial Order STM Deficits in Dyslexia. *Dyslexia*, 18(2), 94-109. <https://doi.org/10.1002/dys.1438>
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1), 593-608.
- Plantinga, J., & Trainor, L. J. (2008). Infants' Memory for Isolated Tones and the Effects of Interference. *Music Perception: An Interdisciplinary Journal*, 26(2), 121-127. <https://doi.org/10.1525/mp.2008.26.2.121>
- Plaza, M., Cohen, H., & Chevrie-Muller, C. (2002). Oral Language Deficits in Dyslexic Children: Weaknesses in Working Memory and Verbal Planning. *Brain and Cognition*, 48(2-3), 505-512. <https://doi.org/10.1006/brcg.2001.1407>

- Roodenrys, S., & Stokes, J. (2001). Serial recall and nonword repetition in reading disabled children. *Reading and Writing, 14*, 379-394. <https://doi.org/10.1023/A:1011123406884>
- Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective Measures of Listening Effort : Effects of Background Noise and Noise Reduction. *Journal of Speech, Language, and Hearing Research, 52*(5), 1230-1240. [https://doi.org/10.1044/1092-4388\(2009/08-0111\)](https://doi.org/10.1044/1092-4388(2009/08-0111))
- Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory, 21*(3), 377-395. <https://doi.org/10.1080/09658211.2012.731070>
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception : Longitudinal evidence from a community-based music program. *Behavioural Brain Research, 291*, 244-252. <https://doi.org/10.1016/j.bbr.2015.05.026>
- Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience, 8*(7), 862-863. <https://doi.org/10.1038/nm1474>
- Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123*(3), 191-201. <https://doi.org/10.1016/j.bandl.2012.09.001>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2021). Auditory and visual short-term memory : Influence of material type, contour, and musical expertise. *Psychological Research*. <https://doi.org/10.1007/s00426-021-01519-0>
- Tijms, J. (2004). Verbal memory and phonological processing in dyslexia. *Journal of Research in Reading, 27*(3), 300-310. <https://doi.org/10.1111/j.1467-9817.2004.00233.x>
- Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia : Evidence from a gating paradigm. *Cortex, 59*, 84-94. <https://doi.org/10.1016/j.cortex.2014.07.012>
- Tillmann, B., Lévêque, Y., Feroni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research, 1640*, 251-263. <https://doi.org/10.1016/j.brainres.2015.10.035>
- Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia : A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition, 71*(3), 259-264. <https://doi.org/10.1016/j.bandc.2009.08.003>
- Wieland, E. A., McAuley, J. D., Dille, L. C., & Chang, S.-E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language, 144*, 26-34. <https://doi.org/10.1016/j.bandl.2015.03.008>
- Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences : Comparing phonological similarity and pitch proximity. *Memory & Cognition, 38*(2), 163-175. <https://doi.org/10.3758/MC.38.2.163>
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnerberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in

- noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225-2234.
<https://doi.org/10.1121/1.4817926>
- Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The Impact of Musicianship on the Cortical Mechanisms Related to Separating Speech from Background Noise. *Journal of Cognitive Neuroscience*, *27*(5), 1044-1059.
https://doi.org/10.1162/jocn_a_00758
- Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110-14115.
<https://doi.org/10.1073/pnas.0504446102>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Foxton, J. M. (2012). Global and local pitch perception in children with developmental dyslexia. *Brain and Language*, *120*(3), 265-270. <https://doi.org/10.1016/j.bandl.2011.12.002>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732-745.
<https://doi.org/10.1111/j.1467-7687.2009.00817.x>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology*, *110*(3), 362-372.
<https://doi.org/10.1016/j.jecp.2011.05.001>

Tables

<i>Measure</i>	<i>Factor(s)</i>	<i>df1</i>	<i>df2</i>	<i>F-value</i>	<i>p-value</i>	η_p^2	<i>W</i>	ϵ		
Response times	d'	Grade	5	86	3.03	0.0144	0.257	-	-	
		Material	1	86	32.329	1.75e - 07	0.273	-	-	
		Grade : Material	5	86	0.675	0.644	0.038	-	-	
		c	Grade	5	86	0.085	0.994	0.013	-	-
		Material	1	86	40.890	8.05e - 09	0.322	-	-	
		Grade : Material	5	86	1.845	0.113	0.097	-	-	
		Grade	5	85	2.338	0.058	0.116	-	-	
		Material	1	85	0.307	0.581	0.004	-	-	
		Type of trial	1	85	0.907	0.344	0.011	-	-	
		Grade : Material	5	85	0.982	0.434	0.055	-	-	
		Grade : Type of trial	5	85	1.554	0.182	0.084	-	-	
		Material : Type of trial	1	85	2.210	0.141	0.025	-	-	
		Grade : Material : Type of trial	5	85	1.675	0.149	0.090	-	-	
	Recency effects	Position	2	261	0.682	0.507	0.008	0.953	-	
		Grade	5	261	1.814	0.118	0.094	-	-	
		Material	1	261	54.850	7.84e - 11	0.387	-	-	
		Position : Grade	10	261	1.030	0.420	0.056	0.953	-	
		Position : Material	2	261	1.551	0.217	0.018	0.911*	0.919	
		(1.84)	(159.83)							
Grade : Material		5	261	1.669	0.151	0.088	-	-		
Position : Grade : Material		10	261	0.872	0.553	0.048	0.911*	0.919		
	(9.19)	(159.83)								

Table 3 : For the easy conditions, results of the repeated measures ANOVAs on each measure (d', c, response times, recency effects). For d' and c, grade was used as between-subject factor (six levels: KG, first, second, third, fourth, fifth grade) and material (two levels: verbal, musical) as within-subject factor. For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and grade as between-subject factors and material as within-subject factor. Significant effects are in bold font. df, degrees of freedom; η_p^2 , partial eta-squared; W, Mauchly's statistic. An asterisk indicates a significant W. If so, corrected degrees of freedom are reported in parenthesis in the df1 and df2 columns and the corresponding statistics are corrected with ϵ : Greenhouse-Geisser estimates of sphericity.

<i>Measure</i>	<i>Factor(s)</i>	<i>df1</i>	<i>df2</i>	<i>F-value</i>	<i>p-value</i>	η_p^2	<i>W</i>	ϵ
d'	Grade	2	47	0.174	0.841	0.029	-	-
	Material	1	47	94.803	7.52e - 13	0.787	-	-

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

Response times	Difficulty	1	47	31.237	1.12e - 06	0.544	-	-
	Grade : Material	2	47	0.462	0.633	0.035	-	-
	Grade : Difficulty	2	47	0.872	0.425	0.062	-	-
	Material : Difficulty	1	47	42.426	4.48e - 08	0.474	-	-
	Grade : Material : Difficulty	2	47	0.125	0.883	0.005	-	-
	c Grade	2	47	0.616	0.545	0.043	-	-
	Material	1	47	92.585	1.09e - 12	0.734	-	-
	Difficulty	1	47	40.558	7.45e - 08	0.432	-	-
	Grade : Material	2	47	2.509	0.0922	0.130	-	-
	Grade : Difficulty	2	47	1.064	0.353	0.038	-	-
	Material : Difficulty	1	47	23.118	1.6e - 05	0.330	-	-
	Grade : Material : Difficulty	2	47	1.775	0.181	0.070	-	-
	Grade	2	47	0.760	0.473	0.031	-	-
	Material	1	47	27.73	3.4e - 06	0.371	-	-
	Difficulty	1	47	3.381	0.072	0.067	-	-
	Type of trial	1	47	1.331	0.254	0.028	-	-
	Grade : Material	2	47	0.314	0.732	0.013	-	-
	Grade : Difficulty	2	47	1.397	0.257	0.056	-	-
	Grade : Type of trial	2	47	0.087	0.916	0.004	-	-
	Material : Difficulty	1	47	0.917	0.343	0.019	-	-
	Material : Type of trial	1	47	7.874	0.007	0.143	-	-
	Difficulty : Type of trial	1	47	0.139	0.711	0.003	-	-
	Grade : Material : Difficulty	2	47	0.168	0.32	0.047	-	-
	Grade : Material : Type of trial	2	47	0.853	0.433	0.035	-	-
	Grade : Difficulty : Type of trial	2	47	1.611	0.211	0.064	-	-
	Material : Difficulty : Type of trial	1	47	0.051	0.822	0.001	-	-
	Grade : Material : Difficulty : Type of trial	2	47	0.019	0.981	0	-	-

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

Recency effects	Position	2	144	32.381	1.79e - 11	0.403	0.922	-
	Grade	2	144	0.355	0.703	0.015	-	-
Material	1	144	161.142	5.92e - 17	0.770	-	-	-
Difficulty	1	144	73.882	2.82e - 11	0.606	-	-	-
Position : Grade	4	144	0.571	0.684	0.023	0.922	-	-
Position : Material	2 (1.68)	144 (80.68)	11.053	1.49e - 04	0.187	0.810*	0.840	-
Position : Difficulty	2	144	42.613	5.68e -14	0.470	0.995	-	-
Grade : Material	2	144	2.274	0.114	0.087	-	-	-
Grade : Difficulty	2	144	0.435	0.65	0.018	-	-	-
Material : Difficulty	1	144	99	2.98 e -13	0.673	-	-	-
Position : Grade : Material	4 (3.36)	144 (80.68)	0.862	0.475	0.035	0.810*	0.840	-
Position : Grade : Difficulty	4	144	1.081	0.37	0.043	0.880	-	-
Position : Material : Difficulty	2	144	38.196	6.26e - 13	0.443	0.960	-	-
Grade : Material : Difficulty	2	144	1.494	0.235	0.059	-	-	-
Position : Grade : Material :	4	144	0.882	0.478	0.035	0.960	-	-
Difficulty								

Table 4 : With older children, results of the repeated measures ANOVAs on each measure (d' , c , response times and recency effects). For d' and c , grade was used as between-subject factor (three levels: third, fourth, fifth grade) and material (two levels: Verbal, musical) and difficulty (two levels: Easy, difficult) as within-subject factors. For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and grade as between-subject factors and material and difficulty as within-subject factors. Significant effects are in bold font. df , degrees of freedom; η_p^2 , partial eta-squared; W , Mauchly's statistic. An asterisk indicates a significant W . If so, corrected degrees of freedom are reported in parenthesis in the $df1$ and $df2$ columns and the corresponding statistics are corrected with ϵ : Greenhouse-Geisser estimates of sphericity.

Measure	Factor(s)	df	df2	F-value	p-value	η_p^2	W	ϵ
d'	Group	1	27	13.25	0.00114	0.743	-	-

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

	Material	1	27	89.526	4.59e - 10	0.869	-	-
	Difficulty	1	27	78.887	1.7e - 09	0.693	-	-
	Group:	1	27	4.981	0.0341	0.269	-	-
	Material							
	Group :	1	27	0.142	0.709	0.004	-	-
	Difficulty							
	Material	1	27	52.691	8.3e - 08	0.661	-	-
	:							
	Difficulty							
	Group:	1	27	0.631	0.434	0.023	-	-
	Material :							
	Difficulty							
c	Group	1	27	0.003	0.957	0	-	-
	Material	1	27	29.615	9.31e - 06	0.719	-	-
	Difficulty	1	27	28.635	1.19e - 05	0.454	-	-
	Group:	1	27	1.752	0.197	0.131	-	-
	Material							
	Group:	1	27	0.322	0.575	0.009	-	-
	Difficulty							
	Material	1	27	14.159	0.000826	0.344	-	-
	:							
	Difficulty							
	Group:	1	27	1.752	0.197	0.061	-	-
	Material :							
	Difficulty							
	Group	1	27	4.825	0.037	0.152	-	-
	Material	1	27	9.911	0.004	0.269	-	-
	Difficulty	1	27	0.196	0.661	0.007	-	-
	Type of trial	1	27	0.874	0.358	0.031	-	-
	Group:	1	27	0.618	0.439	0.022	-	-
	Material							
	Group:	1	27	0.383	0.541	0.014	-	-
	Difficulty							
	Group :	1	27	0.120	0.731	0.004	-	-
	Type of trial							
	Material :	1	27	0.130	0.721	0.005	-	-
	Difficulty							
	Material	1	27	8.625	0.007	0.242	-	-
	: Type of trial							

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

	Difficulty	1	27	0.002	0.969	0	-	-
	: Type of trial							
	Group:	1	27	1.316	0.261	0.046	-	-
	Material :							
	Difficulty	1	27	0.086	0.772	0.003	-	-
	Material :							
	Type of trial							
	Group:	1	27	2.019	0.167	0.07	-	-
	Difficulty							
	: Type of trial							
	Material :	1	27	0.670	0.420	0.024	-	-
	Difficulty							
	: Type of trial							
	Group :	1	27	0.137	0.714	0.005	-	-
	Material :							
	Difficulty							
	: Type of trial							
Recency effects	Position	2	84	25.711	8.87e - 11	0.187	0.862	-
	Group	1	84	10.99	0.00254	0.177	-	-
	Material	1	84	124.42	8.18e - 12	0.542	-	-
	Difficulty	1	84	112.23	2.65e - 11	0.304	-	-
	Position :	2	84	0.151	0.860	0.001	0.862	-
	Group							
	Position :	2	84	12.332	8.30e - 06	0.099	0.847	-
	Material							
	Position :	2	84	25.426	1.12e - 10	0.185	0.985	-
	Difficulty							
	Group:	1	84	10.39	0.00321	0.090	-	-
	Material							
	Group:	1	84	0.058	0.812	0	-	-
Difficulty								
Material	1	84	57.729	2.82e - 08	0.256	-	-	
:								
Difficulty								
Position :	2	84	1.626	0.199	0.014	0.847	-	
Group:								
Material								

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

Position :	2	84	1.023	0.361	0.009	0.985	-
Group:							
Difficulty							
Position :	2	84	13.656	2.53 - 06	0.109	0.973	-
Material							
:							
Difficulty							
Group:	1	84	0.338	0.565	0.002	-	-
Material :							
Difficulty							
Position :	2	84	0.218	0.804	0.002	0.973	-
Group:							
Material :							
Difficulty							

Table 5 : Comparison of children and adult data, results of the repeated measures ANOVAs on each measure (d' , c , response times, recency effects). Ford'andc, group was used as between-subject factor (two levels: fifth grade and adults) and material (two levels: Verbal, musical) and difficulty (two levels: Easy, difficult) as within-subject factors. For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and group as between-subject factors and material and difficulty as within-subject factors. Significant effects are in bold font. df , degrees of freedom; η_p^2 , partial eta-squared; W , Mauchly's statistic. An asterisk indicates a significant W : Mauchly's statistic. An asterisk indicates a significant W . If so, corrected degrees of freedom are reported in parenthesis in the $df1$ and $df2$ columns and the corresponding statistics are corrected with ϵ : Greenhouse-Geisser estimates of sphericity.

Supplementary materials

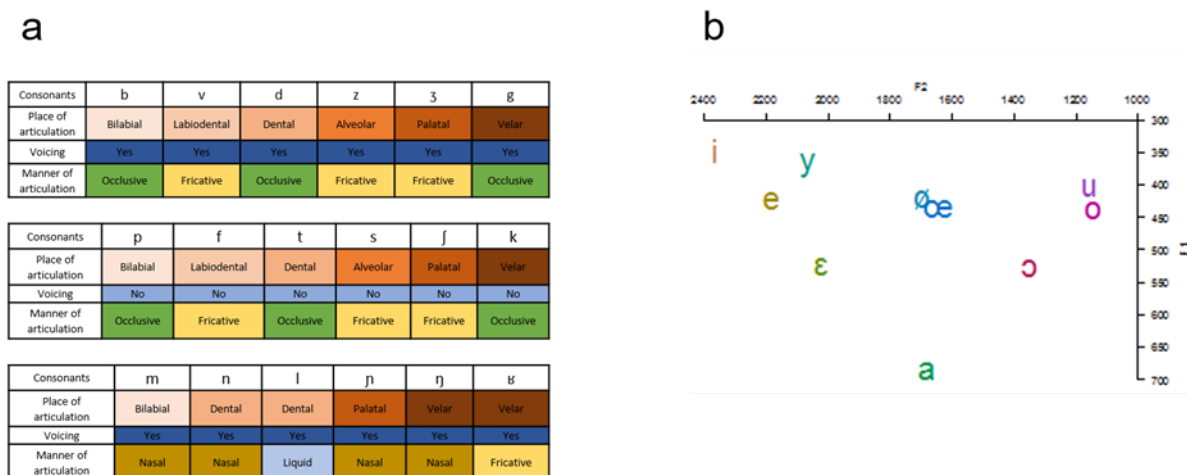


Figure S1 : Choice of consonants and vowels for the syllable sequences. (a): a perceptual distance-table of all French consonants two-by-two (based on Munot & Nève, 2002) was created taking into account voicing, place, and manner of articulation. Six consonants were selected (/f/ /t/ /z/ /g/ /m/ /l/) as they differed by at least two articulatory traits two by two. (b): six vowels were selected as a function of their distance based on their formant values. F1 and F2 formant values were obtained by Georgetown et al. (2012) from a radiophonic corpus gathered by Gendrot & Adda-Decker (2005). We projected all these vowels on a F1/F2 orthonormal system, thus reproducing a quantitative vowel triangle and calculated the Euclidian distance between all vowels. This procedure allowed us to select the vowels with the biggest formant distance two by two: /i/ /e/ /a/ /y/ /ø/ /u/. With the 6 selected consonants and the 6 vowels, 36 syllables were generated and then spoken by a professional mezzo-soprano singer who was given the instruction of keeping pitch as constant as possible. We then selected the clearest and best articulated syllables. Using Adobe Audition (version 3.0), all syllables were cut to obtain sound files of 500 ms and we applied a 20 ms linear fade-out at the end of each file. With the aim to construct isochronous syllable sequences, we extracted the vowels' attacks, allowing us to calculate their perceptual center (p-center, Hidalgo et al., 2017). We then added silence at the beginning of each syllable to line-up all the syllables' p-centers. Finally, we extracted all fundamental frequencies F0 (Praat software, version 6.0.39) of the syllables. They presented a low variability, with a range below one semi-tone (202-212 Hz), i.e., below the smallest frequency difference between tones in the musical task (E2-F2, one semi-tone, see below), so we considered that no equalization of F0 was necessary. Syllables' intensities were equalized based on root mean square (RMS).

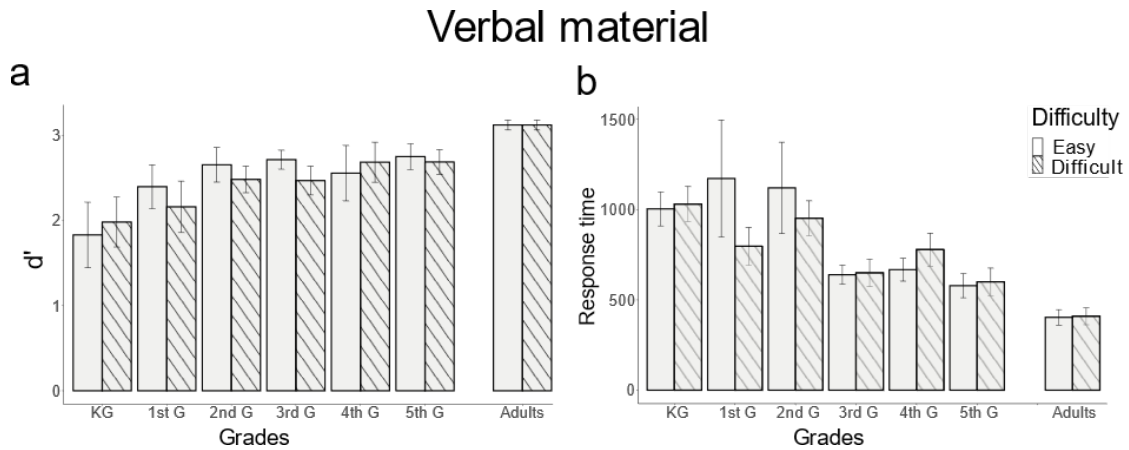


Figure S2 : Performance for the verbal conditions as a function of the difficulty (easy or difficult) for all children’s grades (KG: kindergarten, G: grade) and for adults. a: mean and standard error of children and adult’s sensitivity (calculated as the d'). b: mean and standard error of children and adult’s response time (time in millisecond that subjects spent before giving a “same” or “different” answer).

Two 6x2 mixed-design analyses of variance (ANOVA) were performed for the verbal conditions on d' (Figure S2) and c (Table S1Table 2), with Grade as a between-subjects factor (kindergarten to 5th grade) and with Difficulty as within-subject factor (easy/difficult) as well as one 6x2x2 mixed-design ANOVA for response times analysis (Figure S2) with the aforementioned factors and with type of trial as within-subjects factor (same/different). A 6x2x3 mixed design ANOVA was performed on the percentage of correct responses for different trials, adding the Position factor (2nd, 3rd and 4th position) as within-subject factor (complete results are shown in table S1).

d' . There was a marginally significant effect of Grade in the predicted direction $F(5,86) = 2.105$, $p = .07$, $\eta^2 = .247$. Kindergarten children showed a significantly poorer performance than 3rd to 5th graders ($p < .049$) and a marginally significant poorer performance than 2nd graders ($p = .066$). Other comparisons were not significant ($p > .51$).

Correct response times. We observed a significant effect of Grade $F(5,85) = 2.988$, $p = .016$, $\eta^2 = .149$ with Kindergarten, 1st and 2nd grade children having marginally longer response times than 5th grade children ($p < .058$).

Bias. Comparison of the mean c parameter per grade and condition to 0 showed that only 4th grade children displayed a significantly higher-than-zero c in the difficult condition (Table 2 2). No significant effect was observed in the ANOVA (Table S1).

Recency effects. We observed an effect of Position $F(2,261) = 6.322$, $p = .002$, $\eta^2 = .068$, with a tendency for performance on the 2nd position to be lower than for the 3rd position ($p = .11$) and the 4th position ($p = .098$). Mauchly's test indicated that the assumption of sphericity had been violated for the Position and Difficulty interaction (Mauchly's $W = 0.836$, $p < .001$), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.859$). This interaction was marginally significant $F(1.72,149.49) = 2.742$, $p = .076$, $\eta^2 = .031$, but none of the post-hoc tests were significant ($p > .24$). Other effects mirrored the d' analysis.

When children were gathered in three groups of age (KG-1st, 2nd-3rd, 4th-5th), a main effect of Grade was observed for the d' analysis $F(2,89) = 4.478$, $p = .014$, $\eta^2 = .091$ with the younger group (KG-1st Grade) showing poorer performance than the two older groups ($p < .001$).

<i>Measure</i>	<i>Factor(s)</i>	<i>df1</i>	<i>df2</i>	<i>F-value</i>	<i>p-value</i>	η_p^2	<i>W</i>	ϵ
d'	Grade	5	86	2.105	0.0725	0.247	-	-
	Difficulty	1	86	0.866	0.355	0.010	-	-
	Grade : Difficulty	5	86	0.538	0.747	0.030	-	-
c	Grade	5	86	1.294	0.274	0.191	-	-
	Difficulty	1	86	0.263	0.610	0.003	-	-
	Grade : Difficulty	5	86	1.483	0.204	0.079	-	-
Response times	Grade	5	85	2.988	0.016	0.149	-	-

	Difficulty	1	85	0.826	0.366	0.010	-	-
	Type of trial	1	85	2.592	0.111	0.030		
	Grade : Difficulty	5	85	1.284	0.278	0.070	-	-
	Grade : Type of trial	5	85	0.752	0.587	0.042		
	Difficulty : Type of trial	1	85	0.996	0.321	0.012		
	Grade : Difficulty : Type of trial	5	85	1.025	0.408	0.057		
Recency effects	Position	2	261	6.322	0.002	0.068	0.958	-
	Grade	5	261	2.062	0.078	0.106	-	-
	Difficulty	1	261	0.904	0.344	0.010	-	-
	Position : Grade	10	261	0.570	0.837	0.032	0.958	-
	Position : Difficulty	2	261	2.742	0.076	0.031	0.836*	0.859
		(1.72)	(149.49)					
	Grade : Difficulty	5	261	0.409	0.841	0.023	-	-
	Position : Grade : Difficulty	10	261	0.389	0.934	0.022	0.836*	0.859
	(8.59)	(149.49)						

Table S1 : For the verbal conditions, results of the repeated measures ANOVAs on each measure (d' , c , Response times) with Grade as between-subject factor (6 levels: KG, 1st, 2nd, 3rd, 4th, 5th grade) and Difficulty (2 levels: easy, difficult) as within-subject factor. The analysis of Recency effects was done on the percentage of correct responses for different trials with Position (3 levels: 2nd, 3rd, 4th position) and Grade as between-subject factors and Difficulty as within-subject factor. Significant effects are in bold font. df : degrees of freedom, η_p^2 : partial eta-squared, W : Mauchly's statistic. An Asterisk indicates a significant W . If so, corrected degrees of freedom are reported in parenthesis in the $df1$ and $df2$ columns and the corresponding statistics are corrected with ϵ : Greenhouse-Geisser estimates of sphericity.

Supplementary sounds can be found at

<https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.13188>

Supplementary Sound 1 The Suppinfo_Sound1_musicSTM_easy.wav file is an example of a S1-S2 “different” trial during the easy musical STM block: two sequences of four tones separated by a two-seconds delay. The second tone in S1, A2, is replaced at the same position in S2 by a tone, C2, that changes the contour of the melody. S1 contour: up-down-up (E2-A2-D2-F2). S2 contour: down-up-up (E2-C2-D2-F2).

Supplementary Sound 2 The Suppinfo_Sound2_musicSTM_difficult.wav file is an example of a S1-S2 “different” trial during the difficult musical STM block: two sequences of four tones separated by a two-seconds delay. The second tone in S1, G2, is replaced at the same position in S2 by a tone, A2,

Section 3: Development of auditory cognition in 5- to 10-year-old children: focus on musical and verbal short-term-memory

that does not change the contour of the melody. S1 contour: up-down-up (E2-G2-D2-F2). S2 contour: up-down-up (E2-A2-D2-F2).

Supplementary Sound 3 The Suppinfo_Sound3_verbalSTM_easy.wav file is an example of a S1-S2 “different” trial during the easy verbal STM block: two sequences of four syllables separated by a two-seconds delay. The second syllable in S1, /mi/, is replaced at the same position in S2 by a syllable that changes by its consonant and its vowel, /lu/.

Supplementary Sound 4 The Suppinfo_Sound4_verbalSTM_difficult.wav file is an example of a S1-S2 “different” trial during the difficult verbal STM block: two sequences of four syllables separated by a two-seconds delay. The fourth syllable in S1, /la/, is replaced at the same position in S2 by a syllable that changes only by its vowel, /lu/.

4 Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

Article available as preprint: Ginzburg, J., Cheylus, A., Collard, E., Ferreri, L., Tillmann, B., Moulin, A., & Caclin, A. (2023). Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study. bioRxiv, 2023-10. <https://doi.org/10.1101/2023.10.19.563108>

4.1 General introduction

As reviewed in the first and third section of this PhD, auditory short-term memory (STM) for musical and verbal material plays a key role in children's development of communication skills. Investigating auditory STM for both musical and verbal material is a pivotal endeavor to understand cognitive development. Investigating the neural correlates of auditory STM for different types of materials can shed light on the shared and distinct mechanisms that support these cognitive processes and help us gain knowledge about which of these mechanisms is impaired in language-related learning disorders. Functional Near-Infrared Spectroscopy (fNIRS) offers a unique opportunity to non-invasively examine the hemodynamic responses associated with auditory STM in the developing brain. By harnessing the advantages of fNIRS, such as its silent nature and relative accessibility for children, we can gain insights into how children process and retain both musical and verbal information. As the exploration of auditory STM with fNIRS is relatively new, in the current study, we first conducted two experiments on healthy adults, aiming at characterizing the engagement of prefrontal regions in auditory STM for musical and verbal material. The first experiment aimed at replicating existing fMRI studies as proof of concept and the second one aimed at precisely characterizing the hemodynamic signatures of prefrontal regions while processing STM for musical and verbal material by manipulating memory load.

4.2 Article

Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

Jérémie Ginzburg¹, Anne Cheylus¹, Elise Collard¹, Laura Ferreri^{2,3}, Barbara Tillmann^{1,4}, Annie Moulin¹, Anne Caclin¹

¹ Université Claude Bernard Lyon 1, INSERM, CNRS, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France

² Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

³ Laboratoire d'Étude des Mécanismes Cognitifs, Université Lumière Lyon 2, Lyon, France

⁴ Laboratory for Research on Learning and Development, LEAD – CNRS UMR5022, Université de Bourgogne, Dijon, France

Running title: fNIRS in musical and verbal short-term memory

Corresponding author: Jérémie Ginzburg, jeremie.ginzburg@inserm.fr

Data and code availability statement: The data and the code that support the findings of this study are available on request from the corresponding author.

CRedit authorship contribution statement: **Jérémie Ginzburg:** Conceptualization, Investigation, Methodology, Visualization, Resources, Software, Writing – original draft, Writing – review & editing. **Anne Cheylus:** Methodology, Software, Writing – review & editing. **Elise Collard:** Investigation, Writing – review & editing. **Laura Ferreri:** Conceptualization, Writing – review & editing, Funding acquisition. **Barbara Tillmann:** Conceptualization, Resources, Writing – review & editing, Funding acquisition. **Annie Moulin:** Conceptualization, Resources, Writing – review & editing, Funding acquisition. **Anne caclin:** Conceptualization, Supervision, Methodology, Resources, Software, Writing – review & editing, Funding acquisition.

Conflict of interest statement: The authors declare no conflict of interest.

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

Acknowledgments: We thank Lilou Spinosi, Solène Houdeline, Laurine Milon, and Julie Robin for their help with data recording, as well as Federico Curzel for helpful discussions and Lesly Fornoni for her help with ethics and participants recruitment.

Funding: This work was conducted within the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) of Université de Lyon, within the program “Investissements d’avenir” (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). This work was funded by a “Pack Ambition Recherche” (COGAUDYS project) from the Auvergne-Rhône-Alpes Region, awarded to A. Caclin, A. Moulin, L. Ferreri, and B. Tillmann.

Abstract

Auditory short-term memory (STM) is a key process in auditory cognition, with evidence for partly distinct networks subtending musical and verbal STM. The delayed matching-to-sample task (DMST) paradigm has been found suitable for comparing musical and verbal STM and for manipulating memory load. In this study, musical and verbal DMSTs were investigated with measures of activity in frontal areas with functional near-infrared spectroscopy (fNIRS): Experiment 1 compared musical and verbal DMSTs with a low-level perception task (that does not entail encoding, retention, or retrieval of information), to identify frontal regions involved in memory processes. Experiment 2 manipulated memory load for musical and verbal materials to uncover frontal brain regions showing parametric changes in activity with load and their potential differences between musical and verbal materials. A FIR model was used to deconvolute fNIRS signals across successive trials without making assumptions with respect to the shape of the hemodynamic response in a DMST. Results revealed the involvement of the dorso-lateral prefrontal cortex (dlPFC) and inferior frontal gyri (IFG), but not of the superior frontal gyri (SFG) in both experiments, in keeping with previously reported neuroimaging data (including fMRI). Experiment 2 demonstrated a parametric variation of activity with memory load in bilateral IFGs during the maintenance period, with opposite directions for musical and verbal materials. Activity in the IFGs increased with memory load for verbal sound sequences, in keeping with previous results with n-back tasks. The decreased activity with memory load observed with musical sequences is discussed in relation to previous research on auditory STM rehearsal strategies. This study highlights fNIRS as a promising tool for investigating musical and verbal STM not only for typical populations,

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

but also for populations with developmental language disorders associated with functional alterations in auditory STM.

Keywords: working memory, fNIRS, DMST, auditory, music, language, memory load

Introduction

Auditory Short-Term Memory (STM) plays a crucial role in the processing of auditory information and enables understanding of a dynamic acoustic environment. STM is the active maintenance of information for a brief duration, typically lasting a few seconds (Baddeley & Hitch, 1974; Cowan, 2008). Auditory STM has been extensively studied for verbal material, but a growing body of research has also studied STM for musical material, revealing both shared and specific mechanisms depending on the type of sounds to maintain (Caclin & Tillmann, 2018). Domain-general mechanisms for processing time-based serial order in auditory sequences have been proposed (Gorin et al., 2016), while domain-specific systems have been suggested for encoding and maintenance of distinct materials (Schulze & Tillmann, 2013). Impairment of verbal STM has been shown to be a hallmark of language-related learning disorders, such as dyslexia (Majerus & Cowan, 2016; Roodenrys & Stokes, 2001; Ziegler et al., 2009, 2011) and Developmental Language Disorder (DLD, Archibald & Gathercole, 2006; Nickisch & Von Kries, 2009; Nithart et al., 2009). A few studies have also observed impairment for musical STM in dyslexic children (Couvignou et al., 2023; Forgeard et al., 2008; Ziegler et al., 2012). Moreover, recent studies have shown a sizeable comorbidity (~30%) between dyslexia and congenital amusia (in both adults: Couvignou et al., 2019, and children: Couvignou & Kolinsky, 2021; Couvignou et al., 2023). Congenital amusia is a lifelong disorder characterized by a specific impairment in musical STM (for reviews, see Peretz, 2016; Tillmann et al., 2015, 2023). Overall, these findings suggest impairment of auditory STM in learning disorders, which could either be domain-general or domain-specific. Together with the various patterns of auditory STM impairments observed in neurological diseases (review in

Caclin & Tillmann, 2018), they stress the importance of exploring auditory STM for both materials (music and speech).

Fronto-temporal networks for musical and verbal STM

Neuroimaging studies of auditory STM have revealed the involvement of distributed brain networks including auditory areas in the superior temporal lobe, frontal areas, such as the dorso-lateral prefrontal cortex (dlPFC) and the inferior frontal gyrus (IFG), parietal areas including the supra-marginal gyrus, as well as other brain regions, notably the cerebellum, basal ganglia, and premotor areas (for a review, see Caclin & Tillmann, 2018). Only a few studies using functional magnetic resonance imaging (fMRI) investigated the brain networks involved in STM for both musical and verbal material. A first study reported overlapping brain networks in the superior temporal, inferior parietal, and frontal areas when participants rehearsed out loud novel melodies and nonsense sentences (Hickok et al., 2003). These findings were supported by a second study that examined sung syllables (Koelsch et al., 2009) in which the authors found similar networks activated during the silent rehearsal of verbal information (syllables) and tonal information (pitches), including mostly lateral frontal regions: left IFG and bilateral pre-motor areas, and a small cluster in the left planum temporale. A subsequent fMRI study with sung syllables compared musicians with nonmusicians and revealed that during encoding, parietal and temporal (auditory) cortices showed significant activation in both groups (Schulze et al., 2011). During rehearsal of verbal or tonal information, activation was observed in lateral prefrontal regions (including the dlPFC and IFG) and subcortical structures. While overlapping for both materials, activations were more extended in frontal areas for musical material in musicians. In a recent fMRI study, Albouy et al (2019) investigated STM for musical and verbal material using a delayed-matching-to-sample task (DMST)

and compared activations with control perceptual tasks in amusic participants and matched control participants. During the maintenance of tonal and verbal information, activations emerged for control participants in left lateral frontal regions including the dlPFC and the IFG when comparing the memory task with the simple perception task. However, no activation was observed in temporal auditory regions, suggesting the specific involvement of lateral frontal regions for the maintenance of musical and verbal material, in line with previous findings (Hickok et al., 2003; Koelsch et al., 2009; Schulze et al., 2011). In contrast to controls, amusics' brain activation patterns revealed several functional alterations during encoding and maintenance of musical information including in the right auditory cortex, the right dlPFC, and the right IFG. These results suggest that networks subtending musical and verbal STM are dissociable when studying a population with a specific deficit for the musical material. Overall, neuroimaging data support the view that temporal regions and lateral frontal regions are involved in the encoding of musical and verbal information, while lateral frontal regions are involved in the maintenance of these information, along with other parietal and subcortical areas in some studies. Moreover, these networks seem to be largely overlapping for both materials (Schulze & Koelsch, 2012) and evidence for specialized brain networks for musical and verbal STM arose in fMRI data only from the comparison of impaired (amusics) or expert (musicians) populations to controls.

Memory load manipulation: n-back tasks

Investigating cognitive effort by manipulating memory load is valuable for understanding the underlying mechanisms of auditory STM. Most neuroimaging research on this topic has done so with the verbal identity n-back task that entails maintenance and manipulation of information, hence working memory (WM) (León-Domínguez et al., 2015). In this task, participants are

presented with a series of stimuli (e.g., letters, digits) and are required to indicate when the current item matches the one presented n trials earlier. The task has the advantage of being applicable in the visual or auditory modality. In a meta-analysis of neuroimaging studies with verbal n -back tasks, Owen et al. (2005) found that, regardless of the modality, activity increases with memory load in frontal regions, including dlPFC, IFG, premotor cortex (PMC), and supplementary motor area (SMA). In a meta-analysis with n -back tasks, Sternberg tasks (Sternberg, 1966), and DMSTs, Rottschy et al. (2012) identified bilateral activation patterns where activity increased with memory load, encompassing frontal regions (dlPFC, IFG, PMC, SMA), the middle cingulate cortex and temporo-occipital areas, regardless of the task, modality, or material (words, pseudo-words, pictures...). Note that in this meta-analysis, all Sternberg tasks and DMSTs used visual material. Although the n -back task is frequently employed in neuroimaging studies to manipulate memory load, it is primarily designed to assess WM rather than STM. Indeed, the n -back task involves actively manipulating and updating information, which is specific of WM processes. Furthermore, neuroimaging studies have demonstrated that distinct lateral frontal regions are engaged in forward digit spans (STM) and backward digit spans (WM) (Owen, 2000; Tian et al., 2014). Therefore, the n -back task may not be the most suitable for specifically exploring the effects of memory load on STM per se, as it involves additional cognitive operations beyond simple storage, maintenance, and retrieval of information. Alternatively, recall tasks can be used to study STM and WM, but verbal serial recall tasks involve specific cognitive operation, as they depend on several phonological, sublexical, lexical and semantic factors (e.g. Allen & Hulme, 2006). Furthermore, serial recall tasks are not easy-to-use for studying musical material due to their reliance on production (Caclin & Tillmann, 2018; but see Gorin et al., 2018;

Williamson et al., 2010 for mixed recall/recognition task adaptations to music). Therefore, DMSTs seem to be specifically suited to investigate memory load in auditory STM. However, they have only rarely been employed in neuroimaging studies manipulating memory load in auditory STM.

Memory load manipulation: DMST

Auditory DMST consists in making a same/different judgement between two sequences separated by a silent delay (Albouy et al., 2019; Caclin & Tillmann, 2018; Ginzburg et al., 2022; Gosselin et al., 2009; Talamini et al., 2021; Tillmann et al., 2009). This paradigm lends itself to memory load manipulation for musical and verbal material, to study both STM and WM (with forward and backward instructions respectively), and has been used with three types of auditory material in behavioral experiments (words, pitch, and timbre sequences: Schulze et al., 2012; Schulze & Tillmann, 2013). Research using electroencephalography (EEG) with auditory DMST has consistently demonstrated that a sustained anterior negativity (SAN) in the electrophysiological signal serves as a reliable index of memory load. As memory load increases, the amplitude of SAN at fronto-central electrodes increases for both pitch and timbre materials (Alunni-Menichini et al., 2014; Guimond et al., 2011; Lefebvre et al., 2013; Lefebvre & Jolicœur, 2016; Nolden et al., 2013). However, due to the limited spatial resolution of EEG, the precise engagement of frontal regions could not be fully explored. Only one neuroimaging study, using magnetoencephalography (MEG), has thus far investigated the specific brain regions showing parametric activation with memory load increase during DMST with musical material (Grimault et al., 2014). By examining MEG data at the end of the silent retention delay, significant correlations between the memory load and the amplitude of the evoked response were reported for bilateral IFGs, bilateral dlPFCs, bilateral auditory cortices, and the right

parahippocampal gyrus. These results indicate that lateral frontal regions (IFG and dlPFC), previously identified to be associated with musical and verbal STM (Albouy et al., 2019; Koelsch et al., 2009; Schulze et al., 2011), show a parametric increase of activity with musical memory load, in keeping with neuroimaging studies using verbal n-back tasks (Owen et al., 2005; Rottschy et al., 2012).

fNIRS studies with memory load manipulation in n-back tasks

In recent years, there has been a growing interest for using functional Near-Infrared Spectroscopy (fNIRS) to study brain functioning. fNIRS is a neuroimaging technique that utilizes light sources and detectors at near-infrared wavelengths to measure changes in cerebral metabolism, which serve as an indirect measure of neuronal activity. When neuronal activity increases in a particular cortical area, a rise in oxygenated hemoglobin (HbO) and a decrease in deoxygenated hemoglobin (HbR) occur concurrently, which can both be recorded with fNIRS. This hemodynamic response peaks around 5 seconds after stimulus onset (Fantini et al., 2018; Huppert et al., 2006; Scholkmann et al., 2014). The temporal dynamics of the fNIRS signal is thus similar to the blood-oxygen level-dependent (BOLD) signal in fMRI. Although having a physiological delay typical of hemodynamic signals and a limited spatial resolution, including the difficulty to record deep brain areas (such as the auditory cortex), fNIRS offers several advantages over other imaging techniques, such as fMRI (Pinti et al., 2020). It provides a quiet and less restrictive environment as it is non-invasive and portable, making it more comfortable and suitable for capturing real-world behavior and associated neural responses in children and clinical populations who may find the fMRI environment distressing (Ferreri et al., 2014). Additionally, fNIRS has a higher tolerance to movement compared to fMRI and electroencephalography (Aslin

& Mehler, 2005). Furthermore, whereas the loud sounds of fMRI can present challenges for studying the auditory modality (and in particular for individuals with language processing difficulties as it requires them to listen in a noisy environment), fNIRS is fully silent (Butler et al., 2020; Hancock et al., 2023).

A number of fNIRS studies have employed verbal n-back task to investigate memory load. In n-back tasks with auditorily or visually presented consonants, activity in bilateral lateral prefrontal cortices (LPFCs, including dlPFC and IFG) increased with increasing memory load (Rovetti et al., 2021), while bilateral medial prefrontal cortices (MPFCs) do not seem to respond to memory load manipulation. These results are in accordance with previous neuroimaging studies showing the involvement of LPFC structures in WM, with activity increasing parametrically with load (Owen et al., 2005; Rottschy et al., 2012). Using major, minor, and dissonant chords as stimuli in a n-back task, Tseng et al. (2018) showed that activity increased parametrically with load in bilateral orbital PFC and IFG. However, the use of complex chords (more than three tones simultaneously) and major/minor/dissonant chords possibly entailed that participants categorized these stimuli based on the abstract representation of consonance/dissonance, rather than pure pitch information. Recording the same brain regions as in Rovetti et al (2021), hearing-aid users and normal-hearing participants displayed an increase of activity with load, leveling at 2-back difficulty in ventral and medial PFCs (Rovetti et al., 2019). This parametric activation was observed for verbal memory load in the auditory and visual modality. To the best of our knowledge, only one study used fNIRS to investigate auditory WM load in children with developmental language disorder (DLD) and typically developing children (TD) using an auditory n-back task with consonants (Hancock et al., 2023). There was an increase of activity with increasing memory load in the left dlPFC (the right dlPFC was

not recorded) and a decrease of activity with increasing memory load in bilateral inferior parietal lobules (IPLs) in the TD group, but not in the DLD group. These results suggest a relationship between DLD and difficulties in engaging neural activity for different auditory WM load in dlPFC and parietal regions. In summary, similarly to fMRI investigations, fNIRS studies using auditory n-back tasks have consistently shown the involvement of lateral frontal regions (IFG and dlPFC) in STM, with activity increasing parametrically with load. In contrast, medial frontal areas appear to have a limited involvement in STM.

Studying auditory STM with Delayed-Matching-to-Sample Tasks in fNIRS

One methodological study has shown that DMST is a suitable paradigm to investigate auditory STM using fNIRS, using only verbal material and a single sequence length (Yamazaki et al., 2020). After listening to (or watching at) a 9-syllable sequence, participants had to maintain the information during a 9-seconds retention delay and compare it to a second 9-syllable sequence that could be either identical or different by one syllable. Within a large array of recording channels over the left frontal and temporal areas, significant activation during the encoding and maintenance phase was observed in the auditory modality in the left IFG and dlPFC respectively, along with other premotor and temporal areas.

The primary objective of the current study was to investigate auditory STM, and more specifically the effect of memory load, for musical and verbal material using a DMST paradigm. Two experiments were conducted to achieve this goal. For both experiments, our focus was on the frontal brain regions consistently reported to be involved in both musical and verbal STM (i.e., IFG and dlPFC). Auditory cortices were not targeted with fNIRS due to their depth

in location. Additionally, we recorded medial frontal regions (SFG) as control, since these areas are not typically activated with the fronto-temporal network involved in auditory STM. Experiment 1 adapted the experimental design of Albouy et al. (2019) comparing musical and verbal DMST to a low-level perception task of equivalent duration. Our aim was to replicate findings obtained with fMRI, namely that the lateral prefrontal regions exhibited stronger activation during the memory task compared to the low-level perception task. In Experiment 2, memory load for musical and verbal material was manipulated by varying sequence length. Overall, the research questions were as follow: (1) Is fNIRS suited to explore frontal activations during auditory DMST? (2) What are the frontal brain regions where activity varies parametrically with memory load? (3) Are these regions similar for musical and verbal material?

Experiment 1

Methods

Participants

Nineteen healthy adults were recruited for Experiment 1. Data from three participants were excluded because of technical problems in the fNIRS signal acquisition. This led to a final sample of sixteen right-handed participants (mean age = 39.2 years, sd = 15.2 years, min = 21 years, max = 62 years, 12 females, mean education level = 14.95 years). They all gave written informed consent to participate in the experiment. Prior to the main experiment, all participants were tested with pure tone audiometry (separately for the two ears, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 6000 Hz, 8000 Hz), the Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al., 2003) and a Pitch Discrimination Threshold (PDT) test (Tillmann et al., 2009). All participants

presented a normal audiometry (hearing threshold lower than 30 dB at any frequency in both ears). No participant presented any pitch perception or pitch memory impairment (MBEA $>$ 25 (maximum score = 30) and PDT $<$ 1 semitone) and they had no or little musical education (mean musical education = 0.1 years sd = 0.5 years). No participant presented any neurological or psychiatric history and none reported any past diagnosis of neurodevelopmental disorder. All study procedures were approved by a national ethics committee (CPP Ile de France VI, ID RDC 2018-A02670-55) and participants received a compensation for their participation. During the first session (~1h30), participants performed the audiometry, MBEA, and PDT. During the second session (~1h15) on a different day, participants underwent the fNIRS testing (see below).

Stimuli construction and task design

Musical and verbal stimuli

Musical and verbal stimuli were the same as in Ginzburg et al. (2022). For musical tasks, six musical tones (created with the software Cubase 5.1 (Steinberg) and a Halion Sampler (Steinberg) using an acoustic piano timbre) belonging to the C major scale were used (C2, E2, G2, B2, D3, F3) with frequencies ranging from 131 to 349 Hz (thus encompassing the fundamental frequency range of the vowel recordings: 202-212 Hz). For verbal tasks, the items were Consonant-Vowel syllables that were selected to show the greatest perceptual distance with each other. Six consonants and six vowels were selected: /f/ /t/ /z/ /g/ /m/ /l/ and /i/ /e/ /a/ /y/ /ø/ /u/, thus resulting into 36 syllables that were recorded by a professional mezzo-soprano singer (for details about syllables construction, see Supplementary Figure S1 in Ginzburg et al., 2022). Six syllables were selected: /fi/ /gu/ /ly/ /mø/ /te/ /za/.

Trial examples can be found at

https://github.com/jeremieginzburg/supp_mat_STM_adults_fNIRS.

Perception and memory tasks

For each trial of perception and memory tasks, participants were asked to listen to two 5-item auditory sequences (S1 and S2; verbal or musical) separated by a silent retention delay of 6000 ms. Each item lasted 500 ms, the silent inter-stimulus interval (ISI) between two items lasted 100 ms. Overall, there was a 600 ms stimulus onset asynchrony (SOA), leading to a duration of 11800 ms for S1-delay-S2. Participants were given 3000 ms to provide a same/different response after the end of S2. The next trial started after a 5000- to 9000-ms randomly-jittered silent interval after the response window. Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used to present stimuli and record responses. Additionally, sixteen 22800- to 26800-ms randomly-jittered silent trials were generated to intersperse within each testing block (see *Procedure* below).

For the memory task, S1 and S2 sequences could be either identical or different. All items were different within a given sequence, and all musical sequences included at least one ascending interval and one descending interval. When S2 was different, a new item could appear equiprobably at the 2nd, 3rd, 4th position (positions 1 and 5 were not used for changes to minimize primacy and recency effects). For each material, six S1 sequences were used to create “same” trials, with S2 identical to S1. Six other S1 sequences were used to create “different” trials, with S2 sequences differing from S1 by one item. For the musical material, the new item in S2 always changed the contour of the sequence (the contour is the up-and-down scheme of a melody). So, if S1 had a down-up-**down**-up contour (e.g. E2-C2-**B2-G2**-F3), S2 could have a down-up-**up**-up contour (e.g. E2-C2-**B2-D3**-F3).

For the perception task, new sequences were created and S1 and S2 were always different. In the S2 sequence, the last two items could either be identical or different. Except when S2 contained two identical last items, all items were different within a given sequence, and for the musical material all sequences included at least one ascending interval and one descending interval. For the musical material, when the last two items were different, these two items could not differ by more than 3 tones. A total of twelve S1-S2 sequences were created for each material, half of them with S2 sequences having two identical last items and half of them with S2 sequences having two different last items. For this perception task, the participants were asked to ignore S1 and to answer whether the two last items of S2 were the same or different.

fNIRS montage and data acquisition

The absorption of near-infrared light was measured at 760- and 850-nm wavelengths at a sampling frequency of 7.81 Hz using a continuous-wave NIRScout device (NIRx Medical Technologies, LLC). The data were collected using the NIRStar 15.3 acquisition software. Eight light sources and twelve light detectors were attached to a cap with a 10-20-system marking for probe placement. Additionally, eight 8-mm short-distance channels (one for each source) recorded systemic signal.

The montage was created using fOLD (fNIRS Optodes' Location Decider, Zimeo Morais et al., 2018), which allows placement of optodes in the international 10-20 system to maximize coverage of chosen anatomical regions as defined by one of five segmentation atlases. For the segmentation atlas, we chose the AAL2 (Automated Anatomical Labeling, Rolls et al., 2015) to generate a montage (Figure 1) covering the inferior frontal gyri (IFG) and dorsolateral prefrontal cortex (dlPFC) as they were shown with fMRI to be involved in the encoding and retention phase in memory recognition tasks

(Albouy et al., 2019). As a control, our montage additionally covered medially the superior frontal gyri (SFG), as they have not been shown to be involved in memory recognition tasks.

The final montage included 22 measurement channels with distances ranging from 2.9 to 4.1 centimeters (the montage is represented on Figure 1 and see Supplementary Table S1 for maximum recording specificity of each channel). Data from two channels that had the highest specificity for the precentral gyrus were not analyzed as they were not in the scope of the current study. We discarded two midline channels because they presented noisy signal. Overall, we retained 18 channels for analysis. Among those channels and as calculated by the fOLD software, in each hemisphere, three of them had the highest level of recording specificity for IFG and four of them for dlPFC. Four channels had the highest specificity for bilateral SFGs.

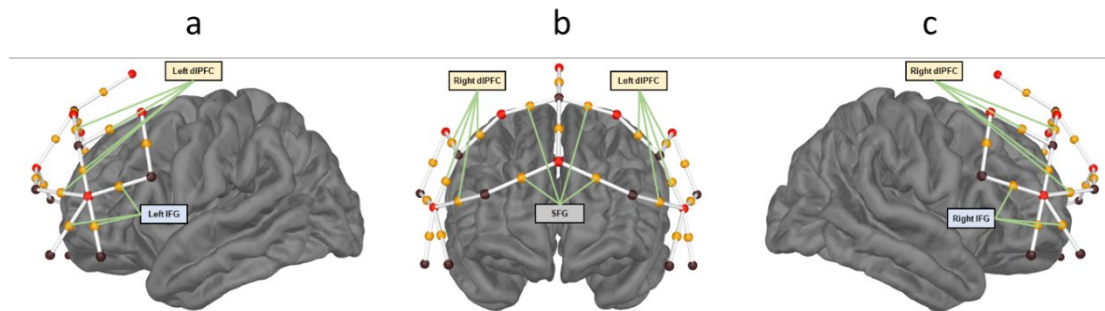


Figure 1: Location of sources (black spheres) and detectors (red spheres) and their corresponding midpoint channels (white lines with orange spheres) for (a) left; (b) front; and (c) right views of the brain. Green lines illustrate the ROIs used for analysis: 3 channels for each IFG (lateral views), 4 channels for each dlPFC (front and lateral views), 4 channels covering bilateral SFGs (front view).

Procedure

The experimenter measured the participant's head circumference to determine cap size. Cap alignment was verified and adjusted if needed so that the probe at Cz was located halfway in the nasion-to-inion and the tragus-to-tragus measurements. The participant was then led into a dimly lit, sound-attenuated booth where, if needed, the participant's hair was moved around

the optode locations using a thin wooden stick to provide clear access to the scalp. Optodes were then placed according to the pre-established montage. fNIRS signal was calibrated, checked for quality, and optode placement on the scalp was readjusted until a satisfying signal quality was obtained before proceeding. Right before testing, the experimenter gave the task instructions to the participant.

Presentation of stimuli and recording of behavioral responses was controlled by Presentation[®] software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, <http://www.neurobs.com>). Triggers were sent from Presentation[®] to the fNIRS acquisition system using a parallel port. Participants sat approximately 50 cm from the computer screen, auditory stimuli were played through Sennheiser HD-250Pro headphones at a comfortable intensity (80 dB SPL A-weighted measured with a Larson-Davis System 824 with an AEC101 IEC 318 Artificial Ear Coupler). Headphones were positioned behind the optodes (primarily located on frontal area) and were sufficiently flexible to avoid any interference with the optical fibers. Participants gave their response with a computer mouse with their right-hand (left-click for ‘same’ response and right-click for ‘different’ response).

All participants underwent four blocks of sixteen trials, one block per task and per material: musical perception, musical memory, verbal perception, verbal memory. Among the sixteen trials, twelve of them were stimulation trials as described above (for trial examples, see Figure 2a). Half of the trials were “same” trials and half of the trials were “different” trials. Stimulation trials were pseudo-randomly presented with the constraint that no more than three consecutive ‘same’ (or ‘different’) trials could be presented. Four “silent” trials were played on the 1st, 6th, 11th, and 16th trial position of a block in order to allow for the hemodynamic signal to return to baseline (Balters et al., 2021)

and to serve as an implicit baseline in the Finite Impulse Response (FIR, see below) analysis (Cairo et al., 2004; Kharitonova et al., 2015). Each of the four blocks lasted around 6 minutes and was preceded by a four-trial training block with feedback using the same task and material. For the test blocks, no feedback was given. The order of test blocks was counterbalanced across participants using a Latin square balancing for first-order carryover effects using the *crossdes* R package (Sailer, 2022).

Behavioral data analysis

Measures of d-prime (d') and criterion (c) were obtained according to Signal Detection Theory (SDT) for each task, material, and participant (Macmillan & Creelman, 2005). Hit corresponded to a correct answer for a different trial. False alarm corresponded to an incorrect answer for a same trial. d' , or sensitivity, was calculated using the *psycho* package (Makowski, 2018) as the z-score of False Alarms subtracted from the z-score of Hits. The criterion was calculated as the mean z-score of Hits and False-alarm rates multiplied by minus one and reflect an observer's bias to say yes (in our case "different") or no ("same"), an unbiased observer having a value around 0. A liberal bias (tendency to say "different") results in a negative c , a conservative one results in positive c . Correction of extreme values was made following Hautus (1995) who recommends the use of a log-linear rule that consists of increasing each cell frequency of the contingency table by 0.5, irrespective of the content of each cell. Furthermore, we analyzed the response times (RT) of participants after the end of S2 for correct trials only, averaged separately for each task, material, and each type of trial (same/different). Note that participants had 3 seconds to answer, otherwise the response was counted as irrelevant and thus not considered in the behavioral analysis. Those missed trials represented 0.01 % of all trials.

Analyses were conducted using Bayesian Statistics that allow the direct comparison of the predictions of several hypotheses (including the null model) and to estimate a degree of logical support or belief regarding effects of interest and their interactions (Wagenmakers et al., 2018). We report Bayes Factor (BF_{10}) as a relative measure of evidence of an effect compared to the null model. Traditionally, a BF_{10} between 1 and 3 is considered as weak evidence for the tested model, between 3 and 10 as positive evidence, between 10 and 100 as strong evidence and higher than 100 as decisive evidence. Similarly, to interpret the strength of evidence in favor of the null model, a BF_{10} between 0.33 and 1 is considered as weak evidence, a BF between 0.01 and 0.33 as positive evidence, a BF between 0.001 and 0.01 as strong evidence and a BF lower than 0.001 as decisive evidence (Lee & Wagenmakers, 2014). For clarity purposes, we report information about the best model only.

Using the R *BayesFactor* package (Morey & Rouder, 2022), d' and c were submitted to a Bayesian repeated-measure ANOVA including task (two levels: perception and memory), material (two levels: musical and verbal), and their interaction as fixed factors. Overall, four models were tested (task, material, task + material, task + material + task:material) and compared to the null model. As recommended by Van Den Bergh et al. (2022), participants were added to all models as random factors using the *lmBF* function of the *BayesFactor* package. Paired Bayesian t-tests were performed as post-hoc tests if the best model included the interaction. Correct RTs were submitted to the same Bayesian ANOVA, with the addition of the type of trial (same/different) factor.

Additionally, one-sample Bayesian t-tests against 0 were performed on the criterion for each task and material.

Furthermore, we report the results of the analysis of effects using the *bayesfactor_inclusion* function from the R from the *bayestestR* package (Makowski et al., 2019) that compares between models that do or do not incorporate a specific effect, such as a factor or an interaction. The resulting measure, $BF_{inclusion}$, serves as a relative indicator of the evidence favoring the inclusion of a factor.

fNIRS data pre-processing and signal deconvolution

fNIRS data were pre-processed using the NIRS Brain-AnalyzIR toolbox (Santosa et al., 2018) and custom-written scripts. Raw intensity signals were first converted to changes in optical density. To correct for motion artifacts from excessive head movements, we applied Temporal Derivative Distribution Repair (TDDR) (Fishburn et al., 2019), a robust regression approach to remove large fluctuations in the optical density signal (motion artifacts), while keeping smaller fluctuations (hemodynamic activity). Corrected optical density were then band-pass filtered between 0.01 and 0.2 Hz to remove cardiac (~ 1.2 Hz) and respiratory activity (~ 0.25 Hz). Finally, corrected and filtered optical densities were transformed into (de)oxygenated hemoglobin concentrations using the modified Beer-Lambert Law.

Data were then processed with a General Linear Model (GLM) using a Finite Impulse Response (FIR) model to deconvolute the signals from successive trials for each channel. This method does not make assumptions regarding the shape of the hemodynamic response and allows for an unconstrained estimation of the full hemodynamic response during stimulation and maintenance. To do so, 36 one-second boxcar regressors were fitted around S1 onsets to encompass the total duration of stimulation (-5 seconds to 30 seconds around S1 onset) for each task and material. A boxcar regressor per block (encompassing the entire twelve stimulation trials and four silence trials) was added to account for

possible HbO/HbR signal changes across blocks of recordings. Finally, data from all short-channels (eight HbO and eight HbR measures) were orthogonalized and added as regressors of no-interest in the GLM in order to further clean the signal from systemic components (Luke et al., 2021). Overall, for each chromophore (HbO/HbR), each recording channel of each participant was regressed using an Ordinary Least Squares (OLS) GLM with a design matrix including block regressors, 36 1-second boxcar regressors for each task and material (thus 144 regressors for the FIR models), and all short-channel signal (16 regressors). Note that no baseline correction is needed in the FIR GLM approach as silent trials, for which no regressors are fitted in the GLM, act as implicit baseline (Cairo et al., 2004; Kharitonova et al., 2015).

fNIRS data analyses

Deconvoluted data were then analyzed using a Bayesian ANOVA on beta coefficients for each 1-s time window of the FIR models and each ROI (see below). The use of Bayes factors to analyze time-course data shows great promise with robustness to type I errors without the need for corrections (see Teichmann et al., 2021 for an empirical comparison between cluster-based corrected time-course data and the Bayes Factor approach). Moreover, as compared to the traditional frequentist approach to time-course data that usually only allows for the comparison of two conditions when applying corrections for multiple testing, Bayesian statistics allow testing for multiple factors and their interaction at each time sample.

Five ROIs (see Figure 1 and Supplementary Table S1) were created with the channels showing the largest specificity for the left or right IFG (3 channels in each hemisphere), the left or right dlPFC (4 channels in each hemisphere), the bilateral SFGs (4 channels). For each participant, time point, task, and material, betas were averaged across channels making up each ROI. These ROI-

averaged betas were then tested across participants, for each time point from -5 to 18 seconds around S1 onset (i.e., until 6 seconds after the end of S2, as the deconvoluted signal would then include the motor response which we did not intend to analyze) for each ROI with a repeated measure Bayesian ANOVA, as for behavioral data, including task (two levels: perception and memory), material (two levels: musical and verbal), and their interaction as fixed factors. Participants and their interaction with the task and material factors (random slopes) were added to all models as random factors. We report here the best model (as compared to the null model, $BF_{10} > 1$) for each time sample. Paired Bayesian t-tests were performed as post-hoc tests if the best model included the interaction.

We report in the main text only HbO results, as HbO tends to show higher amplitude changes and a higher SNR (Pinti et al., 2020) than HbR. However, HbR results are available in supplementary materials and commented in the discussion.

Results

Behavioral results

For d' (Figure 2b), the best model explaining the data included both fixed factors (task and material) and their interaction ($BF_{10} = 3.5e+5$). The analysis of effects across matched models revealed decisive evidence for the task ($BF_{inclusion} = 9.76e+03$) and the material ($BF_{inclusion} = 278.57$) effects and weak evidence for the task:material interaction ($BF_{inclusion} = 1.06$). d' was higher for the verbal material as compared to the musical material and higher for the perception task as compared to the memory task (Figure 2b). Post-hoc Bayesian t-tests revealed strong evidence for the task effect in the musical material ($BF_{10} = 39.46$) and positive evidence for the task effect in the verbal material ($BF_{10} = 8.33$).

For correct RTs (Figure 2c), the best model was the one including only the task effect, but with only weak evidence relative to the null model ($BF_{10} = 2.8$). RTs were slightly longer in the Perception task compared to the Memory task.

For c , results are reported in supplementary Figure S1.

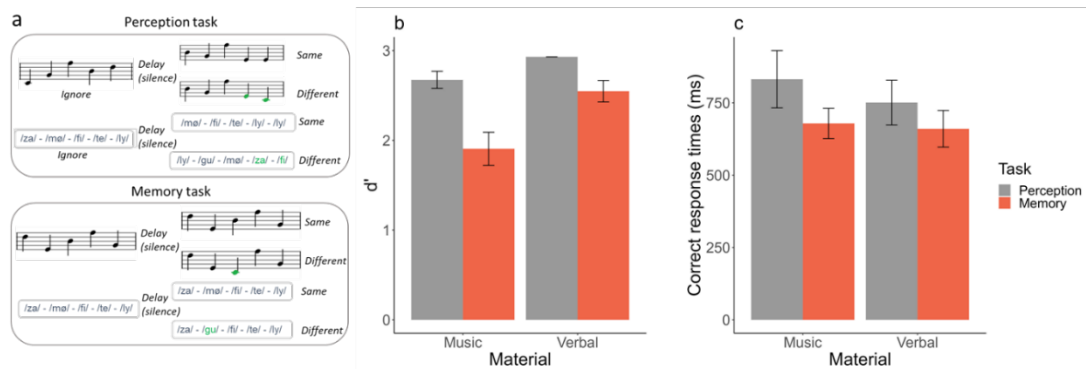


Figure 2: Tasks and average performance as a function of task (perception in grey /memory in orange) and material (musical/verbal). (a) Examples of trials for the perception and memory tasks for musical and verbal materials. (b) Mean and standard error of participant's sensitivity (d'). (c) Mean and standard error of averaged response times for correct trials (time in millisecond that participants spent after the end of S2 before giving a "same" or "different" answer).

fNIRS results

HbO deconvoluted signals within the targeted ROIs are represented with corresponding statistics in Figure 3 (topographic representations are available in supplementary Figure S2 and HbR results in supplementary Figure S3).

For the left IFG, we found weak evidence from -1 to 1 seconds around S1 onset for the model including the task effect ($1.47 < BF_{10} < 1.84$) and weak evidence at 8 seconds for the model with the task effect ($BF_{10} = 1.03$), with in both cases higher betas for the memory task than for the perception task.

For the right IFG, we found weak to positive evidence from -3 to 4 seconds around S1 onset for the model including the task effect ($1.11 < BF_{10} < 7$) with higher betas for the memory task than for the perception task.

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

For the left dlPFC, we only found weak evidence at 5 and 6 seconds for the model including the material effect ($1 < BF_{10} < 1.08$), with higher betas for the musical material than for the verbal material.

For the right dlPFC, we found weak to decisive evidence for the model including the task effect from 8 to 17 seconds around S1 onset ($1.07 < BF_{10} < 2.9e+03$) with higher betas for the memory task than the perception task. At the 14 second post-S1 time sample, we found strong evidence for the model including all effects and their interaction ($BF_{10} = 56.5$). Post-hoc tests at this time sample revealed strong evidence for the task effect only for the verbal material ($BF_{10} = 27$) with higher betas for the memory task than the perception task.

For the SFG, no model was better than the null model to explain the data (all $BF_{10} < 0.76$).

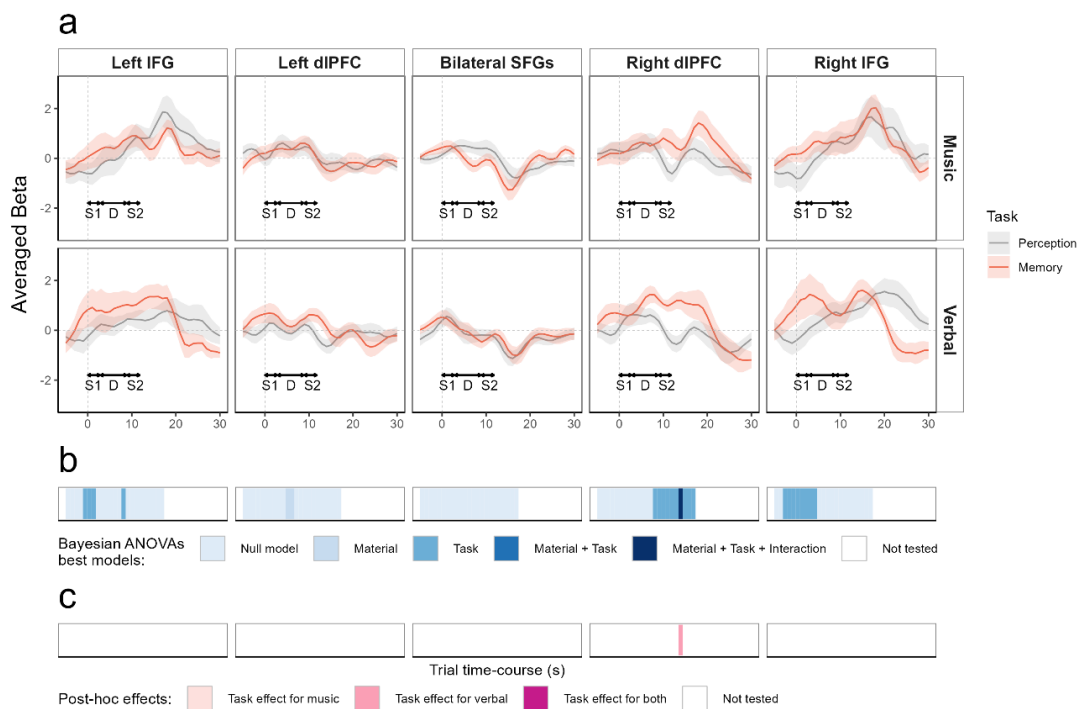


Figure 3 : (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbO data for all participants ($n=16$), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time window ranging from -5 to 30 seconds around S1 onset (grey dotted vertical line), for the perception task (grey) and the memory task (red). Top panel: musical material; bottom panel: verbal material. For clarity purposes, S1, D: silent retention delay,

and S2 durations are indicated with double-headed arrows. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared to the null model ($BF_{10} > 1$). No statistics were performed after 18 seconds (end of S2 + peak of the hemodynamic response, $\sim 6s$), because the deconvoluted signal would then include the motor response which we did not intend to analyze. (c) time-course representation of Bayesian post-hoc analysis performed only if the best model included the interaction. Purple shades indicate if there was evidence in favor of the task effect ($BF_{10} > 1$) for the musical material, the verbal material, or both.

Discussion

Experiment 1 showed better behavioral performance and faster response times for the verbal material, as compared to the musical material for both perception and memory tasks. These results are in accordance with a previous study testing children with the same stimuli and the same memory paradigm that reported higher performance for the verbal material than for the musical material (Ginzburg et al., 2022). Other studies using different verbal stimuli, but the same memory task, reported a material effect in the opposite direction as observed here (verbal performance $<$ musical performance, Albouy et al., 2019; Tillmann et al., 2009). This is due to the use of other verbal stimuli than the current study: mono-syllabic words, resynthesized to obtain a constant fundamental frequency and differing only by consonants in order to exploit the phonological similarity effect (Baddeley, 1966). In contrast, our verbal material aimed for phonologically distinct items as they differ in vowel as well (hence, easier perception and more contrasting material to be memorized). As expected and as in Albouy et al (2019), we also observed higher performance in the perception task compared to the memory task for both materials. Analysis of criterion revealed a weak tendency for a conservative bias for the musical material (more errors in ‘different’ trials) and a more liberal bias for the verbal material (more errors in ‘same’ trials). Based on these results, we increased the sequence length of the verbal material as compared to the musical material for Experiment 2, in order to equalize performance between materials.

HbO deconvoluted fNIRS signals over frontal areas showed a clear pattern of activation in lateral frontal channels over the time-course of the trial for both materials and both tasks (see supplementary Figure S2), with a higher increase in activity for the memory tasks as compared to the perception tasks. Conversely, a deactivation takes place in channels placed above medial frontal areas but does not seem to be different for the memory and the perception tasks. Analyses of activation in frontal ROIs showed evidence for higher activation in the memory task as compared to the perception task in the right dlPFC for both materials between 8 and 17 seconds after S1 onset ($BF_{10} > 1$) and in particular 10 to 13 seconds after S1 onset (all $BF_{10} > 107$). Considering that the hemodynamic response peaks ~5-6 seconds after stimulus onset (Pinti et al., 2020), signals for these time samples should mainly correspond to the cortical activity during the silent retention delay (2.9 – 8.9 s). In bilateral IFGs, we also observed higher activation in the memory task compared to the perception task between -1 and 1 seconds for the left hemisphere ($1.5 < BF_{10} < 1.8$) and between -3 to 4 seconds in the right hemisphere ($3.9 < BF_{10} < 7$). Given the hemodynamic delay, these responses in IFGs correspond to an anticipation period (before 0 s). The activation observed during the anticipation period in this experiment is likely attributable to the block design employed, where each block comprised trials of the same task and material. As in depth encoding of S1 is only needed in memory blocks, participants seemed to have adopted different attending strategies depending on the blocks. To avoid such block-related anticipatory effects, conditions were randomized within blocks in Experiment 2.

The analysis on HbR deconvoluted signal (supplementary Figure S3) mirrored HbO results in lateral frontal regions: we observed in the right dlPFC evidence for a stronger decrease of HbR for the memory task as compared to

the perception task between 9 and 17 seconds after S1 onset ($BF_{10} > 1$) and in particular, 12 to 14 seconds after S1 onset (all $BF_{10} > 13.4$). We also observed a stronger decrease of HbR for the memory task compared to the perception task in the right IFG between -1 to 6 seconds ($1 < BF_{10} < 114$). Finally, no effect of task was observed for the SFG. Overall, these results are in accordance with previously reported findings that observed the involvement of lateral PFC areas in auditory memory processes, but not of medial frontal areas (Owen et al., 2005; Rottschy et al., 2012; Rovetti et al., 2021). They thus confirm that fNIRS is a suitable technique to explore the involvement of frontal areas in musical and verbal STM processes using a DMST paradigm.

Experiment 2

Methods

Participants

Twenty-four healthy adults were recruited for Experiment 2 (mean age = 28.9 years, sd = 9.3 years, min = 21 years, max = 66 years, 6 left handed, 16 females, mean education level = 15.44 years, mean musical education = 0.48 years, sd = 1 year). Participants underwent the same inclusion procedure as for Experiment 1. They all gave written informed consent to participate in the experiment (ethical authorization: CPP Ile de France VI, ID RDC 2018-A02670-55) and were given a compensation for their participation. The first session (~ 1h30) was the same as for Experiment 1 and the second session during which participants underwent fNIRS recordings lasted around 1h30.

Stimuli and task design

In Experiment 2, only the STM task was used. There were three memory load (ML) levels that differed in sequence length: ML1, ML2, and ML3. For the musical material, the three MLs consisted in respectively, four-, five-, and

six-item sequences; for the verbal material, they consisted in respectively six-, seven- and eight-item sequences. Within a trial, S1 and S2 always had the same number of items. Item duration, inter-stimulus interval, delay duration, response time-window and jittered inter-trial interval were as described for Experiment 1. Due to the ML manipulation, the durations of S1-delay-S2 were respectively 10600 ms, 11800 ms and 13000 ms for ML1, ML2, and ML3 for the musical material, and respectively 13000 ms, 14200 ms and 15400 ms for ML1, ML2, and ML3 for the verbal material.

Twelve S1 sequences were created for each ML level and for each material: six S1 sequences were used to create “same” trials, with S2 identical to S1. Six other S1 sequences were used to create “different” trials, with S2 sequences differing from S1. Only six items per material were available in order to maximize phonological discriminability and as the verbal stimuli are intended to be used in children with language disorders as well as in adults. Since there were sequences with more than six items, we allowed for item repetitions within a given sequence with the following constraints: the first item could not be repeated within a sequence, items could be repeated only with at least 2 items in-between, one item could not be repeated more than three times, and three- or four-item patterns could not be repeated. When S2 was different, two adjacent items were switched (instead of introducing a new item) thus systematically changing S2 contour for the musical material. Any item could be switched with the next one, except for the first item. When it was possible, there was an equiprobable number of sequences with each position of item switch (e.g., for musical ML1 “different” S2 sequences, there were three trials with a switch between the 2nd and the 3rd item and three trials with a switch between the 3rd and the 4th item). When it was not possible, the remaining number of trials was randomly assigned to an item position switch (e.g. for

verbal ML1 “different” S2 sequences, there were five trials with an item-switch respectively between the 2nd, the 3rd, the 4th and the 5th item with the adjacent one and one randomly assigned trial with a switch between the 3rd and the 4th items). Additionally, twenty-four 26400- to 30400-ms randomly jittered silent trials were generated to intersperse within each testing block.

fNIRS data acquisition and montage

Data acquisition and fNIRS montage were the same as described in Experiment 1.

Procedure

Participants underwent six blocks of sixteen trials each. Each block could randomly contain any stimulation trial as there were only memory trials for Experiment 2 and we wanted to avoid condition-dependent anticipation effects (as observed in Experiment 1 with a block design). There could be no more than three consecutive “same” (or “different”) trials, no more than three consecutive trials of the same memory load, and no more than four consecutive trials of the same material. Four silent trials were displayed on the 1st, 6th, 11th and 16th trial position of each block. There was a six-trial training block at the beginning of the experiment, with one trial from each ML condition, half of them “same” and half of them “different”. Then, participants underwent the six consecutive test blocks. Each of the six blocks lasted around 6 minutes. The order of test blocks was counterbalanced across participants using a Latin square balanced for first-order carryover effects using the *crossdes* R package (Sailer, 2022).

Behavioral data analysis

Behavioral analysis was performed, as for Experiment 1, on d' , c , and correct RTs. For d' and c , a Bayesian repeated-measure ANOVA was performed

with Memory Load (three levels: ML1, ML2, ML3), Material (two levels: musical and verbal), their interaction as fixed factors and participants as random factor. Overall, four models were tested (memory load, material, memory load + material, memory load + material + memory load:material) and compared to the null model. Paired Bayesian t-tests were performed as post-hoc tests if the best model included the memory load effect or the interaction. Correct RTs were submitted to the same Bayesian ANOVA, with the addition of the type of trial (same/different) as factor. Additionally, one-sample Bayesian t-tests against 0 were performed on the criterion (c) for each memory load and material.

fNIRS data pre-processing and deconvolution

The raw data were preprocessed using the same pipeline as for Experiment 1.

Data were then analyzed with the same FIR model as in Experiment 1. However, as S1 sequences had different durations according to memory load, we centered the deconvolution around the onset of the delay. Hence, 36 one-second boxcar regressors were fitted around delay onset (-9 seconds to 27 seconds). Overall, each recording channel of each participant was regressed using an Ordinary Least Squares (OLS) GLM with a design matrix including block regressors, 36 boxcar regressors around delay onsets for each memory load and material for the FIR deconvolution (216 regressors in total), and all short-channel signals (16 regressors).

fNIRS data analysis

The same ROIs as in Experiment 1 were used, with betas averaged across channels making up each ROI. These ROI-averaged betas were then tested across participants every second from -9 to 12 seconds around delay onset (i.e., until 6 seconds after the end of S2 for the highest ML) in each ROI with a

repeated-measure Bayesian ANOVA including Memory Load (three levels: ML1, ML2, and ML3), Material (two levels: musical and verbal), and their interaction as fixed factors, with the exact same procedure as for the analysis of the behavioral data. Participants were added to all models as random factors with their random slopes for task and material. If the interaction effect was included in the best model, a Bayesian repeated-measure ANOVA with memory load as fixed factor and participants as random factor was performed separately for each material. Then for time samples for which the best model included the memory load effect for a given material, paired Bayesian t-tests were performed as post-hoc tests between each memory load levels (ML1 vs. ML2, ML1 vs. ML3, and ML2 vs. ML3).

As for Experiment 1, we report in the main text only results for HbO, and HbR results are available in supplementary materials (Figure S4) and commented in the discussion.

Results

Behavioral results

For d' (Figure 4a), the best model explaining the data included both fixed factors (memory load and material, decisive evidence, $BF_{10} = 6.1e+15$) but not the interaction between both. The analysis of effects across matched models revealed decisive evidence for memory load ($BF_{inclusion} = 2.83e+08$) and material ($BF_{inclusion} = 1.43e+10$) factors. d' was lower for the verbal material as compared to the musical material, and decreased as load increased (Figure 4a). Post-hoc Bayesian t-tests for the memory load factor revealed decisive evidence for a difference between ML1 and ML2 ($BF_{10} = 3e+04$) with lower d' for ML2 as compared to ML1, decisive evidence for a difference between ML1 and ML3 ($BF_{10} = 2.2e+05$) with lower d' for ML3 as compared to ML1, and weak

evidence for a difference between ML2 and ML3 ($BF_{10} = 2$) with lower d' for ML3 as compared to ML2.

For correct RTs (Figure 4b), the best model was, as for d' , the one including the fixed factors memory load and material (decisive evidence, $BF_{10} = 1.4e+03$). The analysis of effects across matched models revealed decisive evidence for the memory load factor ($BF_{inclusion} = 2.3e+05$) and strong evidence for the material factor ($BF_{inclusion} = 65$). RTs were longer for the verbal material as compared to the musical material, and increased with load (Figure 4b). Post-hoc Bayesian t-tests for the memory load factor revealed positive evidence for a difference between ML1 and ML2 ($BF_{10} = 8.8$) with longer RTs for ML2 as compared to ML1, decisive evidence for a difference between ML1 and ML3 ($BF_{10} = 1.3e+04$) with longer RTs for ML3 as compared to ML1, and weak evidence for a difference between ML2 and ML3 ($BF_{10} = 1$) with slightly longer RTs for ML3 as compared to ML2.

For c, results are reported in supplementary Figure S4.

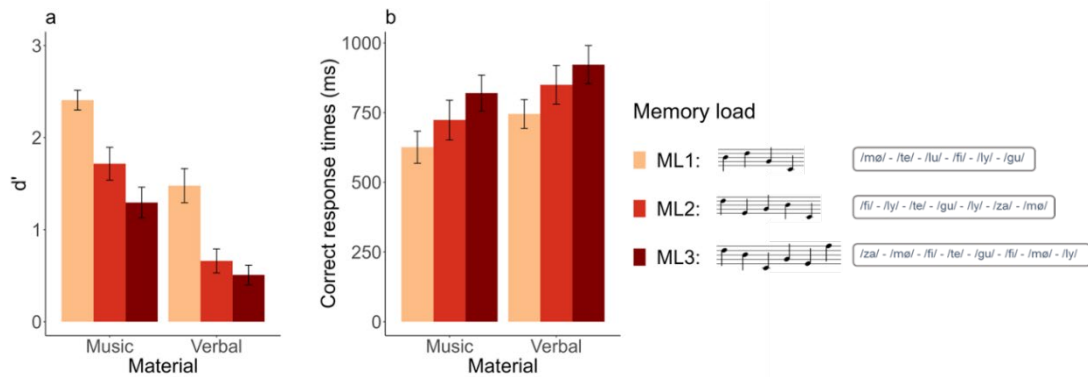


Figure 4: Average performance as a function of memory load (ML) (ML1 in orange/ML2 in red/ML3 in dark red) and material (musical/verbal). (a) Mean and standard error of participants' sensitivity (d'). (b) Mean and standard error of averaged response times for correct trials (time in millisecond that participants spent after the end of S2 before giving a "same" or "different" answer). Examples of S1 sequences for each ML are shown in the legend.

fNIRS results

HbO deconvoluted signals within the targeted ROIs are shown with corresponding statistics in Figure 5 (topographic representations are available in supplementary Figure S5 and HbR results in supplementary Figure S6). For completeness we report all effects below, but our interest was only in models including the memory load factor or its interaction with the material factor.

For the left IFG, we found weak to strong evidence for the model with the material factor at time samples -9, -8, -4, and -3 seconds before delay onset ($1.2 < \text{BF}_{10} < 18.6$) with higher betas for the musical material as compared to the verbal one, and weak to strong evidence for the model with the material factor 8 to 12 seconds after delay onset ($2 < \text{BF}_{10} < 39.4$) with higher betas for the verbal material as compared to the musical one. Importantly, we found positive to strong evidence for the model including the interaction between material and memory load 4 to 7 seconds after delay onset ($3.4 < \text{BF}_{10} < 46.4$). In the musical material, post-hoc tests for the memory load effect revealed a positive to strong evidence for the memory load effect for the four tested time samples ($10.3 < \text{BF}_{10} < 49.7$). Pairwise post-hoc t-tests in the musical material revealed positive to strong evidence for a difference between ML1 and ML2 ($4.4 < \text{BF}_{10} < 10.4$) and weak to positive evidence for a difference between ML1 and ML3 ($1.8 < \text{BF}_{10} < 4.9$) for the four tested time samples. Unexpectedly, in the musical material, betas were higher for the ML1 condition as compared to the ML2 and ML3 conditions. In the verbal material, post-hoc tests for the memory load effect revealed weak to positive evidence for the memory load effect at time samples 4, 5, and 6 seconds ($1.3 < \text{BF}_{10} < 5.3$). Pairwise post-hoc tests in the verbal material revealed weak evidence for a difference between ML1 and ML3 ($1.2 < \text{BF}_{10} < 2.4$) for the three tested time samples and weak evidence for a difference between ML2 and ML3 ($1.2 < \text{BF}_{10} < 1.5$) for time samples 4

and 5. As expected, for the verbal material, betas were higher for the ML3 condition as compared to the ML1 and ML2 conditions. In other words, higher activations were found for the ML1 condition as compared to ML2 and ML3 in the musical material, and higher activations were found for the ML3 condition as compared to ML2 and ML1 in the verbal material.

For the right IFG, we found weak to positive evidence for the model including the material factor -9 to -6 seconds and at 2 seconds around delay onset ($1.1 < \text{BF}_{10} < 10$) with higher betas for the musical material as compared to the verbal one. Importantly, we found weak to strong evidence for the model including the interaction 3 to 6 seconds after delay onset ($1 < \text{BF}_{10} < 19.8$). Post-hoc tests for the memory load effect in the musical material revealed a weak to positive evidence for the memory load effect for time samples 4, 5, and 6 seconds ($2.1 < \text{BF}_{10} < 7.7$). Pairwise post-hoc bayesian t-tests in the musical material and for these three time samples revealed weak to positive evidence for a difference between ML1 and ML2 ($1.2 < \text{BF}_{10} < 3.5$) and weak to positive evidence for a difference between ML1 and ML3 ($2.8 < \text{BF}_{10} < 5$). As for the left IFG for the musical material, betas were higher for the ML1 condition as compared to the ML2 and ML3 conditions. Post-hoc tests for the memory load effect in the verbal material revealed weak to strong evidence for the memory load effect for the four tested time samples ($1.5 < \text{BF}_{10} < 15.7$). Pairwise post-hoc t-tests in the verbal material revealed weak to positive evidence for a difference between ML1 and ML3 ($1.6 < \text{BF}_{10} < 7.8$) for the four tested time samples, and positive evidence for a difference between ML2 and ML3 ($3.9 < \text{BF}_{10} < 4.9$) at time samples 4 and 5 seconds. As expected, for the verbal material, betas were higher for the ML3 condition as compared to the ML1 and ML2 conditions.

For the left dlPFC, we found weak to strong evidence for the model including the material factor at -3 seconds and from 0 to 3 seconds around delay onset ($1 < \text{BF}_{10} < 11.2$) with higher betas for the musical material as compared to the verbal one and weak evidence for the model with the material factor 10 and 11 seconds after delay onset ($1.2 < \text{BF}_{10} < 1.4$) with higher betas for the verbal material as compared to the musical one.

For the right dlPFC, we found weak to strong evidence for the model including the material factor at -9 to -5 seconds and -2 to 3 seconds around delay onset ($1.2 < \text{BF}_{10} < 19.1$) with higher betas for the musical material as compared to the verbal one. We also found weak to strong evidence for the model with the material factor 9 to 12 seconds after delay onset ($1.3 < \text{BF}_{10} < 15.2$) with higher betas for the verbal material as compared to the musical one. We found strong evidence for the model including material and memory load as factors -4 and -3 seconds before delay onset ($20.4 < \text{BF}_{10} < 44.2$) with higher betas for the musical material as compared to the verbal one. Post-hoc tests averaged over materials for the memory load effect (not shown in Figure 5) revealed weak evidence for a difference between ML1 and ML2 ($1.4 < \text{BF}_{10} < 1.9$) and positive to strong evidence for a difference between ML1 and ML3 ($9.8 < \text{BF}_{10} < 10.1$), with higher betas for ML1 as compared to ML2 and ML3.

For the SFG, we found weak to strong evidence for the model including the material factor from 1 to 10 seconds after delay onset ($1.1 < \text{BF}_{10} < 22.8$) with higher betas for the musical material as compared to the verbal material.

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

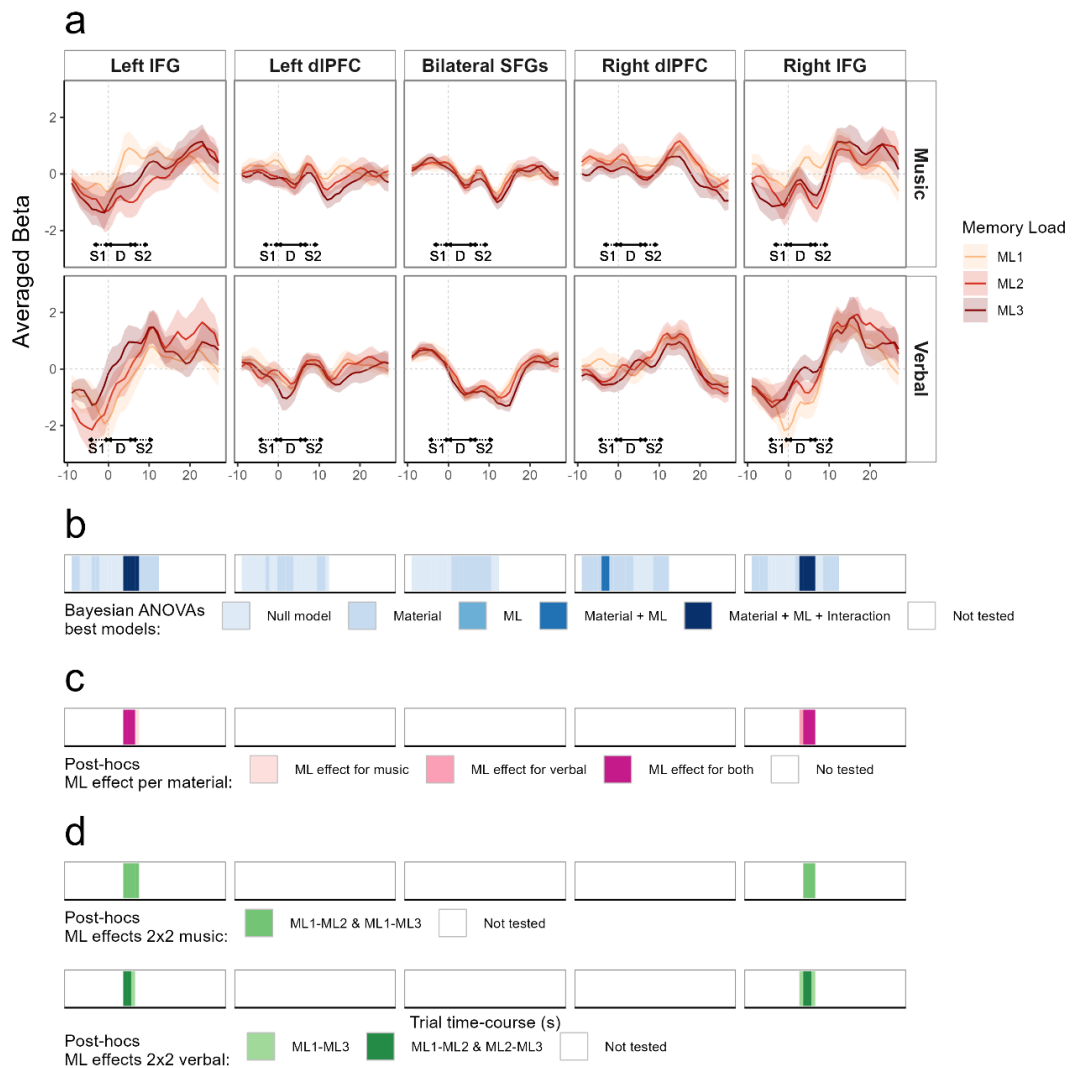


Figure 5 : (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbO data for all participants ($n=24$), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time window ranging from -9 to 27 seconds around delay onset (grey dotted vertical line), for the three memory load levels (ML1/ML2/ML3), for the musical material (top panel) and verbal material (bottom panel). For clarity purposes, S1, silent retention delay (D), and S2 durations are indicated with double-headed arrows, S1 and S2 arrows are dotted to indicate their variable duration according to the memory load. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared to the null model ($BF_{10} > 1$). No statistics were performed beyond 12 seconds after delay onset (end of S2 for the longest sequence + peak of the hemodynamic response, $\sim 6s$). (c) Time-course representation of post-hoc Bayesian ANOVAs per material with memory load (ML) as factor performed only if the best model included the interaction. Purple shades indicate if there was evidence in favor of a memory load effect ($BF_{10} > 1$) for the musical material, the verbal material, or both. (d) Post-hoc Bayesian t -tests comparing memory load levels two by two for the

musical material (top panel) and verbal material (bottom panel) performed only if the previous post-hoc analysis showed evidence for a memory load effect for the corresponding material.

Discussion

The aim of Experiment 2 was to manipulate memory load for musical and verbal STM in order to identify the lateral frontal regions showing parametric involvement with memory load and to explore whether their involvement differs for the two materials.

Behavioral data showed a parametric decrease of performance and an increase of RTs for higher memory loads with strong to decisive evidence for differences between ML1 and the other two memory loads (Figure 4), as previously observed in studies manipulating memory load with DMST (Grimault et al., 2014; Schulze et al., 2012; Schulze & Tillmann, 2013). Analysis of the criterion (c) revealed weak evidence for a liberal bias for low memory loads for the musical material. The performance was overall lower and RTs longer for the verbal material as a consequence of (1) increasing sequence length by two items in the verbal material compared to the musical material and (2) switching from a change of item to a change of order when S2 sequences were different from S1. In future studies using these stimuli, similar performance between musical and verbal material should be reached by increasing sequence length for the verbal material by one item compared to the musical material. Nevertheless, we managed to manipulate memory load in a similar fashion for both materials, as revealed by the absence of interaction between material and memory load for both performance and RTs. These results provide evidence that the DMST paradigm is highly effective in manipulating memory load, in line with previous behavioral studies with auditory material (Schulze et al., 2012; Schulze & Tillmann, 2013) and the few neuroimaging studies employing

DMST for this purpose in the visual (Habeck et al., 2005; Kaiser et al., 2010; Robitaille et al., 2010) and auditory (Grimault et al., 2014) modalities.

As S1 duration was different for each memory load (different sequence lengths), rendering hemodynamic time-course comparison in the response to S1 difficult, we analyzed the hemodynamic responses from the start of the retention delay. Note that material effects in the time ranges corresponding to S1 encoding or S2 processing are likely reflecting the differences in sequence lengths for the two materials. Hence, the interpretation of the data focuses on the delay period. The topographical representation of HbO deconvoluted signal (supplementary Figure S5) showed a parametric modulation of activity in lateral frontal channels for both materials after the delay onset with higher activation for lower memory loads for the musical material and, conversely, higher activation for higher memory loads for the verbal material. In the ROI analysis (Figure 5), the parametric effects in the IFGs were present 3 to 6 seconds after delay onset but not later, confirming that the parametric effect specifically concerned processes at the beginning of the silent retention delay. For the verbal material, we observed the expected activation increase with higher memory loads but intriguingly, for the musical material, the effect of memory load was not the expected one, with higher activation for ML1 as compared to ML2 and ML3. We discuss these results, along with those from Experiment 1, in the following general discussion.

Channels placed along the medial part of the PFC (SFGs) were not sensitive to memory load, as expected. They showed a global deactivation after delay onset regardless of the memory load, but more strongly for verbal stimuli. In the ROI analysis (Figure 5), we observed that the dlPFC and SFG did not show any memory load effect, and that the parametric memory load effects only involved the bilateral IFGs. Furthermore, in the SFG, the material effect

revealed greater deactivation for the verbal material, likely reflecting the increased difficulty of the task, as revealed by the lower performance for the verbal material than for the musical material.

General Discussion

The aims of the present study were three-fold. (1) We aimed at validating the application of fNIRS in investigating frontal activations during auditory DMST. In Experiment 1, the IFG and dlPFC exhibited greater activation for the memory task as compared to the perception task for both materials while the SFG did not exhibit any task effect. (2) Our second goal was to identify the specific frontal brain regions where activity demonstrates parametric variation with memory load. In Experiment 2, only the IFG showed parametric activations with memory load while the dlPFC and the SFG did not respond to memory load manipulation. (3) We wanted to determine whether these regions exhibit similarities when processing musical and verbal materials. Both materials showed similar patterns in Experiment 1 (memory > perception in the IFG and the dlPFC) and memory load processing involved the same region (IFG) in Experiment 2. However, different profiles were observed between materials with increasing activations for increasing memory loads in the verbal material and decreasing activation for increasing memory loads in the musical material.

Involvement of frontal regions in auditory STM

Previous research has consistently indicated the involvement of lateral prefrontal regions during maintenance of musical and verbal material in auditory STM, specifically the IFG and dlPFC. None of the previous studies has reported the involvement of medial prefrontal regions (Grimault et al., 2014; Rottschy et al., 2012). In the current study, both experiments provide converging evidence with these observations. In particular, Experiment 1

provided (1) evidence for the involvement of bilateral IFGs during the anticipation period (-3 to 2 seconds around S1 onset), only for the memory task. These results are an addition to previous fMRI studies that did not investigate anticipation effects. (2) During the maintenance of information (i.e., the silent retention delay), we found that the dlPFC showed higher activation for the memory task as compared to the perception task, while this pattern was not observed in the IFG. The findings in previous studies present inconsistencies regarding the specific involvement of the dlPFC or IFG in the maintenance of musical and verbal STM. While Koelsch et al. (2009) observed activations in the IFG, but not the dlPFC, during the maintenance phase for musical and verbal information, Schulze et al. (2011) and Albouy et al. (2019) reported involvement of both regions during the maintenance period. Given that our study is the first to directly compare materials using a neuroimaging technique that differs from fMRI, combined with the scarcity of existing research in this area, further studies are necessary to draw conclusive insights regarding whether the IFG and/or the dlPFC is preferentially involved in maintenance in auditory STM tasks.

In Experiment 1, our findings revealed, in line with previous fMRI studies, that when comparing memory processes to perception processes in control participants, no difference between musical and verbal material was observed. In addition, we did not observe differences in the medial frontal regions activity for memory processes compared to perception processes. Overall, by demonstrating a differential activation patterns between memory and perception tasks in lateral PFC, our study contributes to the growing body of evidence linking lateral prefrontal regions to the encoding and maintenance of auditory information in STM, a link that was further explored in Experiment 2 by manipulating memory load.

Memory load manipulation for musical and verbal STM

In Experiment 2, our results demonstrated that, as the verbal memory load increased, there was a corresponding increase in activation in bilateral IFGs during the maintenance of information. In contrast, the verbal memory load did not impact bilateral dlPFC and SFG activations. Previous studies in fMRI report a linear increase in activity with visual verbal memory load using n-back tasks (assessing WM) in the left IFG and in bilateral dlPFC (Braver et al., 1997) and in the left IFG and right dlPFC (Cohen et al., 1997). fNIRS studies have also yielded similar results for visual-verbal memory load using n-back tasks (Fishburn et al., 2014; Khaksari et al., 2019). Fishburn et al. (2014) showed an increase in dlPFC with load but not in the IFG, medial PFC, or parietal areas. Khaksari et al (2019) found an increase of activity in bilateral dlPFC memory load (they did not record bilateral IFGs). Interestingly, they also found a parametric increase in medial PFC. However, these results can be mitigated as most of the fNIRS studies analyzed HbO and HbR data, whereas in this study, the authors only analyzed total hemoglobin concentration changes, making the comparison difficult. In the auditory modality, fNIRS studies have found an increase of activity with verbal WM load in bilateral IFGs and dlPFCs (without differentiating them in the fNIRS montage, Rovetti et al., 2021) and in bilateral orbital PFC and IFG, but without recording bilateral dlPFCs (Tseng et al., 2018).

Overall, for verbal material, WM studies consistently report a linear bilateral increase in lateral prefrontal regions with memory load, but most of the time not in medial frontal regions. The relative involvement of the dlPFC and the IFG varies across studies. One potential explanation for the present findings implicating mostly the IFG is that our DMST paradigm targeted STM rather than WM processes. With fMRI, using forward (STM) and backward

(WM) digit span tasks, the orbital part of the IFG showed significant activation for both tasks while the dlPFC showed significant activation only for the WM task. With fNIRS, the contrast between forward and backward digit span tasks shows that only the dlPFC is involved in WM processes (Tian et al., 2014). The relative involvement of IFG, dlPFC, and SFG might thus depend on task requirements, neuroimaging technique, modality, and analysis pipelines.

For musical STM, our study revealed that during the maintenance of information, activity varied parametrically with load in bilateral IFGs but not in bilateral dlPFCs nor SFGs. However, in contrast to the parametric variation for verbal material, we found a parametric decrease of activation with the increase of memory load and more specifically, higher activations for the 1st level of memory load (4-item sequences) as compared to levels 2 and 3 (5- and 6-item sequences). These results are not in line with the only study that investigated memory load effects for musical material using MEG: Grimault et al. (2014) reported an increase of activity with memory load in bilateral IFGs, dlPFCs and temporal regions. The main difference between this latter study and the present one, beyond the use of different neuroimaging techniques, is that they used 4-item sequences as their maximum memory load, when in the present study a 4-item sequence corresponded to the minimum memory load, 6-item sequences being the maximum one. Furthermore, their silent retention interval lasted 2 seconds when ours lasted 6 seconds. Therefore, the task used in the present study was far more challenging than the task used in Grimault et al. (2014). While studies employing verbal n-back tasks have reported deactivation in lateral prefrontal regions when memory loads exceed participants' WM capacity (Nyberg et al., 2009), we can exclude this explanation for the present study. Indeed, our behavioral results demonstrate that participants were able to perform the musical task even at the highest

memory load, and we observed a parametric increase of IFG activity with load in the more difficult verbal task. A possible explanation for these findings could be that when processing musical sequences in the lower memory load condition, participants employ the same strategy as for the verbal material and thus engage the same frontal network, as evidenced by the increased activity in the IFGs in the ML1 condition. However, for higher memory loads, participants might change strategy. It has been shown that for the encoding of verbal information, when the acoustic context facilitates chunking strategies, activity decreases in lateral frontal regions (Ferrerri et al., 2015). Such activity decreases in lateral PFC are also observed when participants are given specific instructions to use a chunking strategy (Matsui et al., 2007). For musical material, participants might have relied on different strategies to handle higher memory loads. A previous study comparing musicians and non-musicians has evidenced the existence of different (and more or less efficient) strategies for musical STM, notably using or not contour information (Talamini et al., 2021). Some of these alternative strategies (chunking, contour-based, etc.) could involve different brain regions than the ones recorded here. Interestingly, when testing the impact of memory load for musical material with MEG, significant clusters of activation were observed not only in frontal areas, but also in temporal and parietal areas (Grimault et al., 2014). Future studies using musical material and a DMST should gather data about the strategy employed by participants to perform the task and record activity in temporal and parietal areas. Moreover, these maintenance strategies could be studied for other auditory material, such as timbre. Indeed, it has been suggested that strategies used to maintain for timbre information in WM differs from tonal and verbal material and that they would rely on sensory imagery rather than internal rehearsal or moto-related processes (Schulze & Tillmann, 2013).

Previous neuroimaging studies comparing musical and verbal STM consistently found overlapping regions involved for both materials, except in impaired or expert populations (Caclin & Tillmann, 2018). Our results support the observation that the same brain regions are recruited for musical and verbal STM, but by manipulating memory load, we found that their involvement differed between the two materials. These results yield great promise for the identification of specific neurophysiological markers of auditory STM for specific materials. Here, by increasing the length of the verbal sequences by two items compared to the musical sequences, we managed to have a similar influence of increases in memory loads on behavior, but with an overall decrease of performance for verbal sequences compared to musical ones. To go further, future studies should equalize performance levels between the two materials, along with recording other regions that have been shown to be involved in memory processes (e.g. temporal and parietal regions). One potential limitation of fNIRS is its limited ability to record deeper brain regions, such as the primary auditory cortex and surrounding areas, which could also be involved in musical memory. While some studies have reported success in detecting auditory cortex activity using fNIRS (Plichta et al., 2011; Santosa et al., 2014), it remains a methodological challenge. Additionally, it would be valuable to investigate the activity of parietal regions, such as the superior and inferior parietal lobule, which activity has been shown to be modulated by memory load for musical and verbal material (Grimault et al., 2014; Rovetti et al., 2021).

Perspectives for studies in children with learning disorders

fNIRS is a promising tool for investigating auditory cognition in children, with typical development and with learning disorders, such as dyslexia and DLD, thanks to its portability, non-invasiveness, and silent nature compared to fMRI. As we have demonstrated in the present study, fNIRS is effective in

measuring neurophysiological markers of auditory STM and can differentiate the effects of memory load between different materials. To our knowledge, only two studies have investigated neural activity in children with developmental language disorder (DLD) with fNIRS so far. Decreased HbO activation in bilateral IFG and parietal regions can be observed in children with DLD during a language comprehension task compared to typically developing (TD) children (Fu et al., 2016) while reduced engagement of the left dlPFC and bilateral IPLs can be found in DLD children during a verbal WM task, compared to TD children. Overall, these previous studies and the current findings highlight the potential of fNIRS as a suitable neuroimaging technique to explore functional differences between children with DLD and TD children, and suggest that variations of cerebral activity in response to memory load manipulation could be an adequate neurophysiological marker. Future studies could investigate further the engagement of lateral frontal regions during auditory STM for musical and verbal material in children with DLD as deficits in STM for both materials have been observed (Couvignou et al., 2023; Couvignou & Kolinsky, 2021; Forgeard et al., 2008; Ziegler et al., 2012). In the long term, a better understanding of the neural mechanisms underlying auditory STM in DLD and their link to the processes reported to be impaired in this population (e.g., syntax processing, phoneme discrimination, etc.) could lead to the development of more effective and targeted interventions for this population.

References

- Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory : FMRI evidence in congenital amusia. *Human Brain Mapping, 40*(3), 855-867. <https://doi.org/10.1002/hbm.24416>
- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*(1), 64-88. <https://doi.org/10.1016/j.jml.2006.02.002>
- Alunni-Menichini, K., Guimond, S., Bermudez, P., Nolden, S., Lefebvre, C., & Jolicoeur, P. (2014). Saturation of auditory short-term memory causes a plateau in the sustained anterior negativity event-related potential. *Brain Research, 1592*, 55-64. <https://doi.org/10.1016/j.brainres.2014.09.047>
- Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*(6), 675-693.
- Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants : Promise, prospects, and challenges. *Journal of Biomedical Optics, 10*(1), 011009. <https://doi.org/10.1117/1.1854672>
- Baddeley, A. D. (1966). The Influence of Acoustic and Semantic Similarity on Long-term Memory for Word Sequences. *Quarterly Journal of Experimental Psychology, 18*(4), 302-309. <https://doi.org/10.1080/14640746608400047>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-90.
- Balters, S., Baker, J. M., Geeseman, J. W., & Reiss, A. L. (2021). A Methodological Review of fNIRS in Driving Research : Relevance to the Future of Autonomous Vehicles. *Frontiers in Human Neuroscience, 15*, 637589. <https://doi.org/10.3389/fnhum.2021.637589>
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage, 5*(1), 49-62. <https://doi.org/10.1006/nimg.1996.0247>
- Butler, L. K., Kiran, S., & Tager-Flusberg, H. (2020). Functional Near-Infrared Spectroscopy in the Study of Speech and Language Impairment Across the Life Span : A Systematic Review. *American Journal of Speech-Language Pathology, 29*(3), 1674-1701. https://doi.org/10.1044/2020_AJSLP-19-00050
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory : Insights from neurodevelopmental and neurological disorders: Musical and verbal short-term memory. *Annals of the New York Academy of Sciences, 1423*(1), 155-165. <https://doi.org/10.1111/nyas.13733>
- Cairo, T. A., Liddle, P. F., Woodward, T. S., & Ngan, E. T. C. (2004). The influence of working memory load on phase specific patterns of cortical activity. *Cognitive Brain Research, 21*(3), 377-387. <https://doi.org/10.1016/j.cogbrainres.2004.06.014>

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604-608.
<https://doi.org/10.1038/386604a0>
- Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia*, *155*, 107811. <https://doi.org/10.1016/j.neuropsychologia.2021.107811>
- Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, *36*(1-2), 1-17.
<https://doi.org/10.1080/02643294.2019.1578205>
- Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology*, 1-47. <https://doi.org/10.1080/09297049.2022.2162031>
- Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research* (Vol. 169, p. 323-338). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Fantini, S., Frederick, B., & Sassaroli, A. (2018). Perspective : Prospects of non-invasive sensing of the human brain with diffuse optical imaging. *APL Photonics*, *3*(11), 110901. <https://doi.org/10.1063/1.5038571>
- Ferreri, L., Bigand, E., Bard, P., & Bugaiska, A. (2015). The Influence of Music on Prefrontal Cortex during Episodic Encoding and Retrieval of Verbal Information : A Multichannel fNIRS Study. *Behavioural Neurology*, *2015*, 1-11. <https://doi.org/10.1155/2015/707625>
- Ferreri, L., Bigand, E., Perrey, S., & Bugaiska, A. (2014). The promise of Near-Infrared Spectroscopy (NIRS) for psychological research : A brief review. *L'Année Psychologique*, *114*(03), 537-569.
<https://doi.org/10.4074/S0003503314003054>
- Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR) : A motion correction method for fNIRS. *NeuroImage*, *184*, 171-179.
<https://doi.org/10.1016/j.neuroimage.2018.09.025>
- Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*.
<https://doi.org/10.3389/fnhum.2014.00076>
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 383-390. <https://doi.org/10.1525/mp.2008.25.4.383>
- Fu, G., Wan, N. J. A., Baker, J. M., Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2016). A Proof of Concept Study of Function-Based Statistical Analysis of fNIRS Data : Syntax Comprehension in Children with Specific Language Impairment Compared to Typically-Developing Controls. *Frontiers in Behavioral Neuroscience*, *10*. <https://doi.org/10.3389/fnbeh.2016.00108>

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

- Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5- to 10-year-old children : Focus on musical and verbal short-term memory. *Developmental Science*, 25(3). <https://doi.org/10.1111/desc.13188>
- Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-Generality of Timing-Based Serial Order Processes in Short-Term Memory : New Insights from Musical and Verbal Domains. *PLOS ONE*, 11(12), e0168699. <https://doi.org/10.1371/journal.pone.0168699>
- Gorin, S., Mengal, P., & Majerus, S. (2018). A comparison of serial order short-term memory effects across verbal and musical domains. *Memory & Cognition*, 46(3), 464-481. <https://doi.org/10.3758/s13421-017-0778-0>
- Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired Memory for Pitch in Congenital Amusia. *Annals of the New York Academy of Sciences*, 1169(1), 270-272. <https://doi.org/10.1111/j.1749-6632.2009.04762.x>
- Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., Zatorre, R., Robitaille, N., & Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch : Evidence from magnetoencephalography. *NeuroImage*, 94, 96-106. <https://doi.org/10.1016/j.neuroimage.2014.03.020>
- Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory : Pitch and ASTM. *Psychophysiology*, 48(11), 1500-1509. <https://doi.org/10.1111/j.1469-8986.2011.01234.x>
- Habeck, C., Rakitin, B. C., Moeller, J., Scarneas, N., Zarahn, E., Brown, T., & Stern, Y. (2005). An event-related fMRI study of the neural networks underlying the encoding, maintenance, and retrieval phase in a delayed-match-to-sample task. *Cognitive Brain Research*, 23(2-3), 207-220. <https://doi.org/10.1016/j.cogbrainres.2004.10.010>
- Hancock, A. S., Warren, C. M., Barrett, T. S., Bolton, D. A. E., & Gillam, R. B. (2023). Functional near-infrared spectroscopy measures of neural activity in children with and without developmental language disorder during a working memory task. *Brain and Behavior*, 13(2). <https://doi.org/10.1002/brb3.2895>
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of cognitive neuroscience*, 15(5), 673-682.
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, 29(2), 368-382. <https://doi.org/10.1016/j.neuroimage.2005.08.065>
- Kaiser, S., Kopka, M.-L., Rentrop, M., Walther, S., Kronmüller, K., Olbrich, R., Weisbrod, M., & Stippich, C. (2010). Maintenance of real objects and their verbal designations in working memory. *Neuroscience Letters*, 469(1), 65-69. <https://doi.org/10.1016/j.neulet.2009.11.045>

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

- Khaksari, K., Condy, E., Millerhagen, J., Anderson, A., Dashtestani, H., & Gandjbakhche, A. (2019). Effects of Performance and Task Duration on Mental Workload during Working Memory Task. *Photonics*, *6*(3), 94. <https://doi.org/10.3390/photonics6030094>
- Kharitonova, M., Winter, W., & Sheridan, M. A. (2015). As Working Memory Grows : A Developmental Account of Neural Bases of Working Memory Capacity in 5- to 8-Year Old Children and Adults. *Journal of Cognitive Neuroscience*, *27*(9), 1775-1788. https://doi.org/10.1162/jocn_a_00824
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory : An fMRI study. *Human Brain Mapping*, *30*(3), 859-873. <https://doi.org/10.1002/hbm.20550>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling : A practical course*. Cambridge university press.
- Lefebvre, C., & Jolicœur, P. (2016). Memory for pure tone sequences without contour. *Brain Research*, *1640*, 222-231. <https://doi.org/10.1016/j.brainres.2016.02.025>
- Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., Zatorre, R. J., & Jolicœur, P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia*, *51*(13), 2939-2952. <https://doi.org/10.1016/j.neuropsychologia.2013.08.003>
- León-Domínguez, U., Martín-Rodríguez, J. F., & León-Carrión, J. (2015). Executive n-back tasks for the neuropsychological assessment of working memory. *Behavioural Brain Research*, *292*, 167-173. <https://doi.org/10.1016/j.bbr.2015.06.002>
- Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, *8*(02). <https://doi.org/10.1117/1.NPh.8.2.025008>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : A user's guide*, 2nd ed. (p. xix, 492). Lawrence Erlbaum Associates Publishers.
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia : The Importance of Serial Order. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.01522>
- Makowski, D. (2018). The psycho Package : An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. <https://doi.org/10.21105/joss.00470>
- Makowski, D., Ben-Shachar, M., & Lüdtke, D. (2019). bayestestR : Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. <https://doi.org/10.21105/joss.01541>
- Matsui, M., Tanaka, K., Yonezawa, M., & Kurachi, M. (2007). Activation of the prefrontal cortex during memory learning : Near-infrared spectroscopy study.

- Psychiatry and Clinical Neurosciences*, 61(1), 31-38.
<https://doi.org/10.1111/j.1440-1819.2007.01607.x>
- Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs*. <https://CRAN.R-project.org/package=BayesFactor>
- Nickisch, A., & Von Kries, R. (2009). Short-Term Memory (STM) Constraints in Children With Specific Language Impairment (SLI) : Are There Differences Between Receptive and Expressive SLI? *Journal of Speech, Language, and Hearing Research*, 52(3), 578-595. [https://doi.org/10.1044/1092-4388\(2008/07-0150\)](https://doi.org/10.1044/1092-4388(2008/07-0150))
- Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading Disabilities in SLI and Dyslexia Result From Distinct Phonological Impairments. *Developmental Neuropsychology*, 34(3), 296-311. <https://doi.org/10.1080/87565640902801841>
- Nolden, S., Bermudez, P., Alunni-Menichini, K., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2013). Electrophysiological correlates of the retention of tones differing in timbre in auditory short-term memory. *Neuropsychologia*, 51(13), 2740-2746. <https://doi.org/10.1016/j.neuropsychologia.2013.09.010>
- Nyberg, L., Dahlin, E., Stigsdotter Neely, A., & Bäckman, L. (2009). Neural correlates of variable working memory load across adult age and skill : Dissociative patterns within the fronto-parietal network. *Scandinavian Journal of Psychology*, 50(1), 41-46. <https://doi.org/10.1111/j.1467-9450.2008.00678.x>
- Owen, A. M. (2000). The role of the lateral frontal cortex in mnemonic processing : The contribution of functional neuroimaging. *Experimental Brain Research*, 133(1), 33-43. <https://doi.org/10.1007/s002210000398>
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm : A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46-59. <https://doi.org/10.1002/hbm.20131>
- Peretz, I. (2016). Neurobiology of Congenital Amusia. *Trends in Cognitive Sciences*, 20(11), 857-867. <https://doi.org/10.1016/j.tics.2016.09.002>
- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders : The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, 999(1), 58-75.
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1), 5-29. <https://doi.org/10.1111/nyas.13948>
- Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., & Fallgatter, A. J. (2011). Auditory cortex activation is modulated by emotion : A functional near-infrared spectroscopy (fNIRS) study. *NeuroImage*, 55(3), 1200-1207. <https://doi.org/10.1016/j.neuroimage.2011.01.011>

- Robitaille, N., Marois, R., Todd, J., Grimault, S., Cheyne, D., & Jolicœur, P. (2010). Distinguishing between lateralized and nonlateralized brain activity associated with visual short-term memory : FMRI, MEG, and EEG evidence from the same observers. *NeuroImage*, *53*(4), 1334-1345. <https://doi.org/10.1016/j.neuroimage.2010.07.027>
- Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*, 1-5. <https://doi.org/10.1016/j.neuroimage.2015.07.075>
- Roodenrys, S., & Stokes, J. (2001). Serial recall and nonword repetition in reading disabled children. *Reading and Writing*, *14*, 379-394. <https://doi.org/10.1023/A:1011123406884>
- Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory : A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830-846. <https://doi.org/10.1016/j.neuroimage.2011.11.050>
- Rovetti, J., Goy, H., Nurgitz, R., & Russo, F. A. (2021). Comparing verbal working memory load in auditory and visual modalities using functional near-infrared spectroscopy. *Behavioural Brain Research*, *402*, 113102. <https://doi.org/10.1016/j.bbr.2020.113102>
- Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional Near-Infrared Spectroscopy as a Measure of Listening Effort in Older Adults Who Use Hearing Aids. *Trends in Hearing*, *23*, 233121651988672. <https://doi.org/10.1177/2331216519886722>
- Sailer, M. O. (2022). *crossdes : Construction of Crossover Designs*. <https://CRAN.R-project.org/package=crossdes>
- Santosa, H., Hong, M. J., & Hong, K.-S. (2014). Lateralization of music processing with noises in the auditory cortex : An fNIRS study. *Frontiers in Behavioral Neuroscience*, *8*. <https://doi.org/10.3389/fnbeh.2014.00418>
- Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms*, *11*(5), 73.
- Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85*, 6-27. <https://doi.org/10.1016/j.neuroimage.2013.05.004>
- Schulze, K., Jay Dowling, W., & Tillmann, B. (2012). Working Memory for Tonal and Atonal Sequences during a Forward and a Backward Recognition Task. *Music Perception*, *29*(3), 255-267. <https://doi.org/10.1525/mp.2012.29.3.255>
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, *1252*(1), 229-236. <https://doi.org/10.1111/j.1749-6632.2012.06447.x>
- Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, *21*(3), 377-395. <https://doi.org/10.1080/09658211.2012.731070>

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping, 32*(5), 771-783. <https://doi.org/10.1002/hbm.21060>
- Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.), 153*(3736), 652-654. <https://doi.org/10.1126/science.153.3736.652>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2021). Auditory and visual short-term memory : Influence of material type, contour, and musical expertise. *Psychological Research*. <https://doi.org/10.1007/s00426-021-01519-0>
- Teichmann, L., Moerel, D., Baker, C., & Grootswagers, T. (2021). *An empirically-driven guide on using Bayes Factors for M/EEG decoding* [Preprint]. Neuroscience. <https://doi.org/10.1101/2021.06.23.449663>
- Tian, F., Yennu, A., Smith-Osborne, A., Gonzalez-Lima, F., North, C. S., & Liu, H. (2014). Prefrontal responses to digit span memory phases in patients with post-traumatic stress disorder (PTSD) : A functional near infrared spectroscopy study. *NeuroImage: Clinical, 4*, 808-819. <https://doi.org/10.1016/j.nicl.2014.05.005>
- Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. In *Handbook of Clinical Neurology* (Vol. 129, p. 589-605). Elsevier. <https://doi.org/10.1016/B978-0-444-62630-1.00033-0>
- Tillmann, B., Graves, J. E., Talamini, F., Leveque, Y., Fornoni, L., Hoarau, C., Pralus, A., Ginzburg, J., Albouy, P., & Caclin, A. (2023). Auditory cortex and beyond : Deficits in congenital amusia. *Hearing Research, 437*, 108855. <https://doi.org/10.1016/j.heares.2023.108855>
- Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia : A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition, 71*(3), 259-264. <https://doi.org/10.1016/j.bandc.2009.08.003>
- Tseng, Y.-L., Lu, C.-F., Wu, S.-M., Shimada, S., Huang, T., & Lu, G.-Y. (2018). A Functional Near-Infrared Spectroscopy Study of State Anxiety and Auditory Working Memory Load. *Frontiers in Human Neuroscience, 12*, 313. <https://doi.org/10.3389/fnhum.2018.00313>
- Van Den Bergh, D., Wagenmakers, E.-J., & Aust, F. (2022). *Bayesian Repeated-Measures ANOVA : An Updated Methodology Implemented in JASP* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/fb8zn>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35-57. <https://doi.org/10.3758/s13423-017-1343-3>
- Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences : Comparing phonological similarity and pitch proximity. *Memory & Cognition, 38*(2), 163-175. <https://doi.org/10.3758/MC.38.2.163>

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

- Yamazaki, H., Kanazawa, Y., & Omori, K. (2020). Advantages of double density alignment of fNIRS optodes to evaluate cortical activities related to phonological short-term memory using NIRS-SPM. *Hearing Research, 395*, 108024. <https://doi.org/10.1016/j.heares.2020.108024>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Foxtton, J. M. (2012). Global and local pitch perception in children with developmental dyslexia. *Brain and Language, 120*(3), 265-270. <https://doi.org/10.1016/j.bandl.2011.12.002>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science, 12*(5), 732-745. <https://doi.org/10.1111/j.1467-7687.2009.00817.x>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology, 110*(3), 362-372. <https://doi.org/10.1016/j.jecp.2011.05.001>
- Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports, 8*(1), 3341. <https://doi.org/10.1038/s41598-018-21716-z>

Supplementary material

fNIRS montage: channel specificity

Hemisphere	Source	Detector	Cortical structure	Specificity (%)
Left	AF7	F5	IFG (p. Triangularis)	54.20
			MFG	25.02
	F7	F5	IFG (p. Triangularis)	82.28
			IFG (p. Orbitalis)	11.66
	FC5	F5	IFG (p. Triangularis)	68.70
			IFG (p. Opercularis)	19.56
	AF3	F5	MFG	66.90
			IFG (p. Triangularis)	22.08
	F3	F1	MFG	68.06
			SFG	30.84
	F3	F5	MFG	60.23
			IFG (p. Triangularis)	38.27
	F3	FC3	MFG	81.08
			PG	9.36
	AF3	AFz	SFG	46.49
			SFG, medial	37.44
	Fz	F1	SFG, medial	40.89
			SFG	39.90
FC5	FC3	Precentral Gyrus	46.40	
		IFG (p. Opercularis)	18.71	
Right	AF8	F6	IFG (p. Triangularis)	38.12
			MFG	37.98
	F8	F6	IFG (p. Triangularis)	73.88
			IFG (p. Orbitalis)	18.76
	FC6	F6	IFG (p. Triangularis)	59.66
			IFG (p. Opercularis)	25.22
	AF4	F6	MFG	69.69
			SFG	14.98
	F4	F2	MFG	65.21
			SFG	31.75
	F4	F6	MFG	59.39
			IFG (p. Triangularis)	39.41
	F4	FC4	MFG	70.58
			IFG (p. Opercularis)	10.29
	AF4	AFz	SFG	42.17
			SFG, medial	36.66
	Fz	F2	SFG, medial	40.68
			SFG	35.24

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

	FC6	FC4	Precentral Gyrus	45.52
			IFG (p. Opercularis)	30.53
Interhemispheric sulcus	Fz	AFz	SFG, medial	43.16
	Fz	FCz	SFG, medial	23.98

Table S1: Each of the 22 recording channels is composed of a source and a detector (located in standard 10-20 positions). We report in the table the two cortical structures for which each channel has the highest specificity (in percent) according to the fOLD software (Zimeo Morais et al., 2018). SFG: Superior Frontal Gyrus, MFG: Middle Frontal Gyrus, IFG: Inferior Frontal Gyrus. The channels forming the IFG, the dlPFC (corresponding to MFG), and the SFG ROIs are highlighted with light blue, light gold, and grey colors respectively. Channels unused in the ROI analysis are not highlighted.

Experiment 1

Behavioral results: criterion

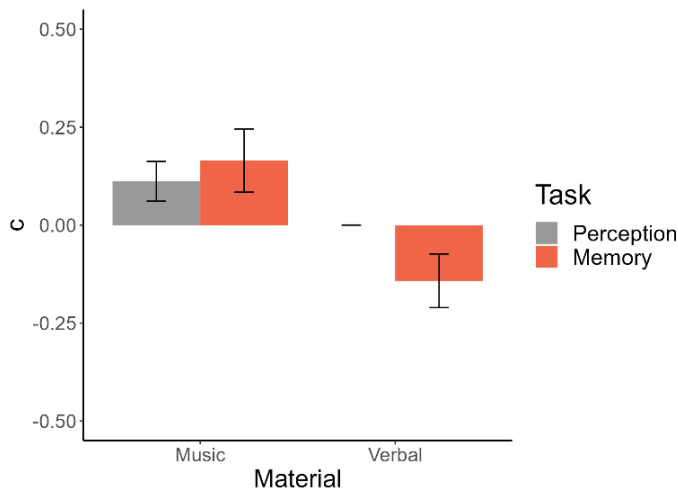


Figure S1: Mean and standard error of criterion (c) as a function of the task (perception in grey/memory in orange) and material (music/verbal).

For the criterion (c), the best model explaining the data included only the material factor (strong evidence, $BF_{10} = 89.2$) with a more conservative criterion for the musical material than for the verbal material. One-sample Bayesian t-tests revealed weak evidence for a difference compared to 0 for the perception and memory tasks in the musical material and for the memory task in the verbal material ($1.3 < BF_{10} < 1.7$). Note that for the perception task in

the verbal material, all participants displayed a criterion of 0 (no bias) because they performed correctly for all materials.

Topographic representation of HbO results

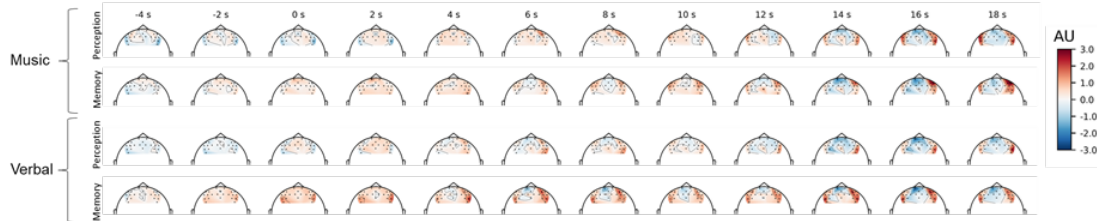


Figure S2: time-course of the topographic representation of the deconvoluted HbO fNIRS signal. Averaged beta across participants are represented every two seconds from -4 to 18 seconds around S1 onset for each channel, each task (perception/memory) and each material (music/verbal). AU: arbitrary units.

HbR results within targeted ROIs

HbR results within the targeted ROIs are summarized in Figure S2.

For the left IFG, we found weak evidence at -5 and -4 seconds before S1 onset for the model including the material effect ($1 < \text{BF}_{10} < 1.6$) higher betas for the verbal material as compared to the musical material.

For the right IFG, we found weak evidence -5 seconds before S1 onset for the model including the material effect ($\text{BF}_{10} = 2.3$) with higher betas for the verbal material as compared to the musical material. We found weak to decisive evidence -1 to 6 seconds around S1 onset for the model including the task effect ($1.4 < \text{BF}_{10} < 160.1$) with lower betas for the memory task as compared to the perception task.

For the left dlPFC, we found weak to strong evidence -5 to 3 seconds around S1 onset for the model including the task effect ($1.1 < \text{BF}_{10} < 28.2$) with lower betas for the memory task as compared to the perception task.

For the right dlPFC, we found weak to strong evidence -2 to 2 seconds and 9 to 17 seconds around S1 onset for the model including the task effect (1.3

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

(< BF_{10} < 13.7) with lower betas for the memory task as compared to the perception task.

For the SFG, we found weak to positive evidence -2 to 1 seconds around S1 onset for the model including the task effect ($1.6 < BF_{10} < 3.3$) with lower betas for the memory task as compared to the perception task.

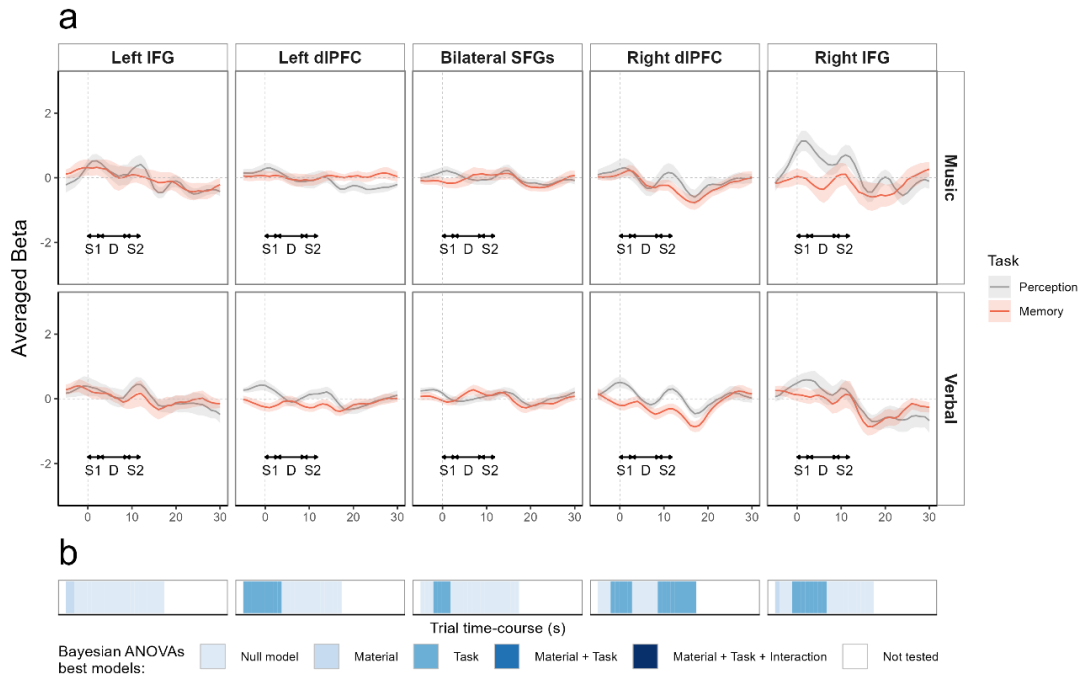


Figure S3 : (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbR data for all participants ($n=16$), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time window ranging from -5 to 30 seconds around S1 onset (grey dotted vertical line), for the perception task (grey) and the memory task (red). Top panel: musical material; bottom panel: verbal material. For clarity purposes, S1, delay (D), and S2 durations are indicated with double-headed arrows. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared to the null model ($BF_{10} > 1$). No statistics were performed after 18 seconds (end of S2 + peak of the hemodynamic response, ~ 6s), because the deconvoluted signal would then include motor responses which we did not intend to analyze.

Experiment 2

Behavioral results: criterion

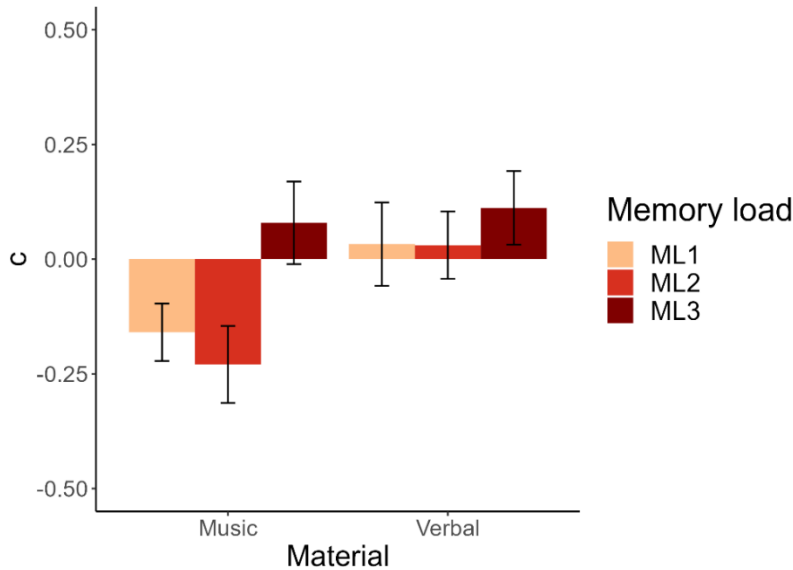


Figure S4: Mean and standard error of criterion (c) as a function of memory load (ML1 in orange/ML2 in red/ML3 in dark red) and material (music/verbal).

For the criterion (c), the best model explaining the data included both the condition and the material factor (positive evidence, $BF_{10} = 6.3$). The analysis of effects across matched models revealed weak evidence for the condition effect ($BF_{inclusion} = 1.72$) and positive evidence for the material effect ($BF_{inclusion} = 3.5$). For the material effect, participants displayed a more conservative criterion for the verbal material as compared to the musical material. Post-hoc Bayesian t-tests for the memory load factor averaged across materials revealed positive evidence for the null model between ML1 and ML2 ($BF_{10} = 0.24$), weak evidence for a difference between ML1 and ML3 ($BF_{10} = 2.5$) with a more conservative criterion for ML3 than for ML1, and positive evidence for a difference between ML2 and ML3 ($BF_{10} = 8$) with a more conservative criterion for ML3 than for ML2. One-sample Bayesian t-tests revealed weak evidence for a difference compared to 0 revealed positive evidence for a difference against for ML1 and ML2 conditions in the musical material (3

$< \text{BF}_{10} < 4.2$). For all other comparisons, weak to positive evidence for the null model (no difference compared to 0) was found ($0.2 < \text{BF}_{10} < 0.5$).

Topographic representation of HbO results

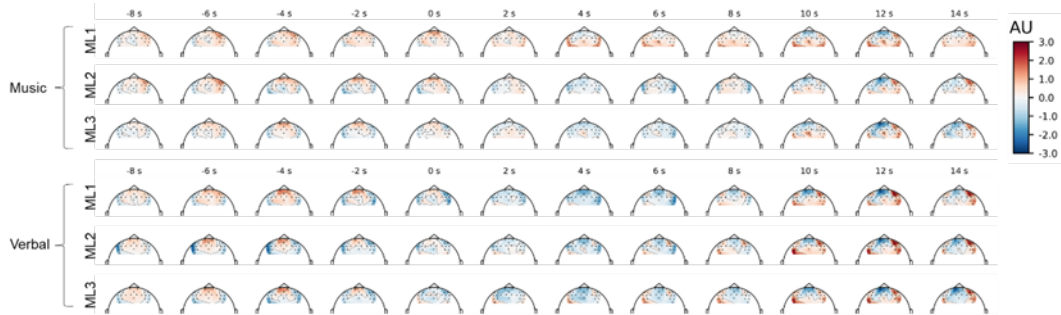


Figure S5: Time-course of the topographic representation of the HbO deconvoluted fNIRS signal. Averaged beta across participants are represented every two seconds from -8 to 14 seconds around delay onset (time sample 0) for each channel, each memory load (ML1/ML2/ML3), and each material (music/verbal). AU: arbitrary units.

HbR results within targeted ROIs

For the left IFG we found weak evidence at -8 seconds before delay onset for the model including the material effect ($\text{BF}_{10} = 1.3$) with lower betas for the musical material as compared to the verbal material. We found also weak to positive evidence at 3 and 4 seconds and 10 to 12 seconds after delay onset for the model including the material effect ($1.1 < \text{BF}_{10} < 4.7$) with lower betas for the verbal material as compared to the musical material.

For the right IFG we found weak evidence at -9 and -8 seconds before delay onset for the model including the material effect ($1 < \text{BF}_{10} < 1.4$) with lower betas for the verbal material as compared to the musical material. We found weak to strong evidence -2 to 4 seconds and 7 to 12 seconds around delay onset for the model including the memory load effect ($1 < \text{BF}_{10} < 28.6$). Post-hoc tests averaged over materials for the memory load effect (not shown in Figure S6) revealed weak to positive evidence 0 to 4 seconds after delay onset for a difference between ML1 and ML2 ($1.4 < \text{BF}_{10} < 7.7$) with lower betas for

the ML2 condition as compared to the ML1 one. We found weak evidence 7 to 12 seconds after delay onset for a difference between ML1 and ML3 ($1.5 < \text{BF}_{10} < 2.8$) with lower betas for the ML3 condition as compared to ML1 condition. Finally, we found weak to positive evidence for all tested time samples (-2 to 4 and 7 to 12 seconds) for a difference between ML2 and ML3 ($2.1 < \text{BF}_{10} < 5.9$) with lower betas for the ML2 condition as compared to the ML3 condition.

For the left dlPFC, we found weak to positive evidence -2 to 2 seconds around delay onset for the model including the material effect ($1.5 < \text{BF}_{10} < 8$) with lower betas for the musical material as compared to the verbal material.

For the right dlPFC, we found weak to strong evidence at -9 seconds, -4 to 0 seconds and 10 to 12 seconds around delay onset for the model including the material effect ($1 < \text{BF}_{10} < 49.6$) with lower betas for the musical material -9 seconds and -4 to 0 seconds as compared to the verbal material and lower betas for the verbal material 10 to 12 seconds after delay onset as compared to the musical material. We found weak evidence -8 and -7 seconds before delay onset for the model including the interaction between material and memory load ($1.5 < \text{BF}_{10} < 1.7$). Post-hoc tests for the memory load effect in the musical material revealed a positive to strong evidence for the memory load effect for the two tested time samples ($15 < \text{BF}_{10} < 24$). Pairwise post-hoc t-tests in the musical material revealed strong evidence for a difference between ML2 and ML3 ($19.3 < \text{BF}_{10} < 22.3$) with lower betas for the ML2 condition as compared to the ML3 condition. Finally, we found weak to positive evidence 1 to 4 seconds after delay onset for the model including the memory load effect ($1.7 < \text{BF}_{10} < 9.2$). Post-hoc tests averaged over materials for the memory load effect (not shown in Figure S6) revealed positive evidence in all tested time samples for a difference between ML1 and ML2 with lower betas for ML2 as compared to the ML1 condition. There was also weak evidence at 1 and 2 seconds for a difference

Section 4: Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

between ML2 and ML3 ($1.1 < BF_{10} < 1.3$) with lower betas for the ML2 condition as compared to the ML3 condition.

For the SFG, we found weak to strong evidence -1 to 10 seconds around delay onset for the model including the material effect ($1 < BF_{10} < 74.9$) with lower betas for the musical material as compared to the verbal material.

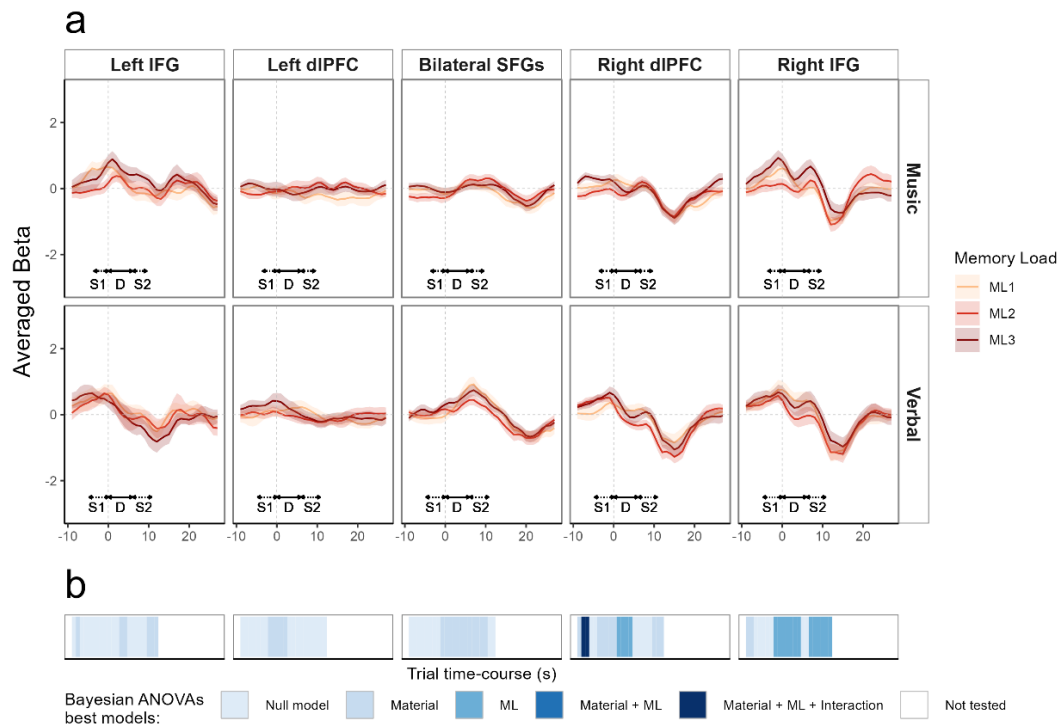


Figure S6 : (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbR data for all participants ($n=24$), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time window ranging from -9 to 27 seconds around delay onset (grey dotted vertical line), for the three memory load levels (ML1/ML2/ML3), for the musical material (top panel) and verbal material (bottom panel). For clarity purposes, S1, silent retention delay (D), and S2 durations are indicated with double-headed arrows, S1 and S2 arrows are dotted to indicate their variable duration according to the memory load. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared to the null model ($BF_{10} > 1$). No statistics were performed beyond 12 seconds after delay onset (end of S2 for the longest sequence + peak of the hemodynamic response, $\sim 6s$).

5 Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

Acknowledgements: We thank all the children who took part in the study, along with their parents. We thank all the study contributors: Anne Cheylus, Annie Moulin, Lucile Lailot, Nathalie Bedoin, Barbara Tillmann, and Julie Robin.

Presentation of preliminary results

5.1 General introduction

As detailed in the previous sections, auditory short-term memory (STM) for musical and verbal material is crucial for the acquisition of communication skills. Indeed, children with learning disorders exhibit impairments in STM for both materials, which can have far-reaching implications for their academic and social well-being. We showed in section 4 that functional near-infrared spectroscopy (fNIRS) is a well-suited neuroimaging tool to explore objective markers of auditory STM in healthy adults. In the present study, we now explore the engagement of prefrontal regions in auditory STM for musical and verbal material in typically developing children and children with learning disorders. We intend to explore the objective markers uncovered in the adult's study to unveil the neural dynamics of the auditory STM impairment that is consistently reported in learning disorders. Data are still being gathered and we will present in this section preliminary results from 11 typically developing children.

5.2 Report on preliminary results

Introduction

Language-related learning disorders, such as dyslexia and developmental language disorders (DLD), have been closely associated with central auditory

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

processing disorders (CAPD), characterized by challenges in effectively processing auditory stimuli within the central auditory system, in the absence of peripheral hearing impairments (Moore et al., 2010). This suggests that dyslexia and DLD impairments stem, at least partly, from deficits in central auditory processes. Developmental dyslexia is a neurodevelopmental disorder that affects around 3–7% of the general population and is characterized by significant and persistent difficulties in acquiring reading skills, despite normal intellectual abilities, no evident sensory or neurological impairments, and adequate educational opportunities (World Health Organization, 2015). Developmental language disorder (DLD, also termed specific language impairment, SLI) is a neurodevelopmental disorder that emerges in early childhood and often continues into adulthood. The prevalence of DLD has been complex to determine due to the constant changes in diagnosis criteria but seems to affect between 6 to 10% of children, making it one of the most prevalent learning disorders (Grimm & Schulz, 2014; Norbury et al., 2016; Tomblin et al., 1997). Individuals with DLD face substantial challenges in learning, understanding, and using spoken language (McGregor, 2020). Both dyslexia and DLD are believed to result from a complex interplay of genetic, environmental, and neural developmental factors. Children with developmental dyslexia or specific subtypes of DLD experience significant reading difficulties, often attributed to underlying phonological challenges (Nithart et al., 2009; Ramus, 2003). The process of learning to read involves grapheme-to-phoneme conversion rules, which demands the ability to perceive, segment, and explicitly manipulate the sounds of spoken words—an aspect commonly known as phonological awareness (Nithart et al., 2009). In this context, it has been hypothesized that these difficulties stem from a deficient phonological STM (Briscoe et al., 2001; Nithart et al., 2009). During reading, the sequence of

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

phonemes must be temporarily stored in STM to be combined and matched with the phonological and lexical representations stored in long-term memory (Nelson & Warrington, 1980). Indeed, STM impairments in dyslexia and DLD have been consistently reported (Forgeard et al., 2008; Majerus & Cowan, 2016; Nithart et al., 2009; Ziegler et al., 2012). In the next paragraphs, we will review current knowledge about auditory STM behavioral and cerebral functional impairments in both disorders (dyslexia and DLD).

Verbal STM in dyslexia and DLD

Verbal STM capacity, measured by digit span or non-word repetition, is typically impaired in both children and adults with dyslexia (Majerus & Cowan, 2016). This deficit may contribute to dyslexia by limiting the simultaneous co-activation of phonological and graphemic information during the reading process, especially during the recoding phase when grapheme-to-phoneme mappings are not yet automatic (Gathercole & Baddeley, 1993). Children with dyslexia consistently score lower than their typically developing (TD) peers in tasks assessing verbal span, forward digit recall, word recall, and nonword repetition (Jeffries & Everatt, 2004; Menghini et al., 2011; Schuchardt et al., 2013). Although the deficit primarily concerns verbal material (Jeffries & Everatt, 2004), there have been suggestions that it could extend to visuo-spatial STM as well (Menghini et al., 2011). This raises the question of whether the verbal STM deficits in dyslexia are attributable to the phonological nature of the items being memorized or reflect more generalized impairments that affect memory regardless of the material being retained.

Majerus & Cowan (2016) have proposed a distinction between item and serial-order processing in verbal STM that would respectively be domain-specific and domain-general. Theoretically, item information (i.e., words in a list), is thought to be encoded by temporarily activating language

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

representations, while serial-order information is commonly believed to depend on distinct processing systems employing temporal, spatial, or magnitude codes (Burgess & Hitch, 1999, 2006; Henson, 1998; Page & Norris, 1998). Developmental studies have shown that serial-order STM capabilities predict lexical and reading progress independently of item STM abilities and are a robust predictor of lexical and reading abilities (Hachmann et al., 2020; Leclercq & Majerus, 2010; Majerus et al., 2006; Martinez Perez et al., 2012; Ordonez Magro et al., 2020). Thus, the distinction between item and serial order STM capacities may be valuable in understanding the nature of verbal STM deficits in dyslexia. If the verbal STM impairments in dyslexia stem solely from phonological processing difficulties, then only item STM performance should be impaired. Alternatively, if additional STM deficits (e.g., for the visuo-spatial modality) are observed, serial-order STM might also be impaired.

Studies in adult with a history of dyslexia have yielded mixed results concerning the potential impairment of item and serial-order STM. Martinez Perez et al. (2013), using tasks aimed at maximizing temporary retention of either item or serial-order information, demonstrated both item and serial-order STM deficits in adults with a history of dyslexia. Moreover, they found that these deficits were statistically independent. Hachmann et al. (2014) observed serial-order STM impairment in both visual and verbal modalities among adults with a history of dyslexia but did not observe verbal item STM impairment. This absence of verbal item STM impairment could possibly be attributed to the strong lexico-semantic component of the verbal items used in their task, as deficits in verbal item STM among dyslexic individuals are believed to arise at the phonological rather than the lexical level (Majerus et al., 2016). Wang et al. (2016) found no indications of either verbal item or serial-order STM impairment among undergraduate university students with self-reported

dyslexia diagnoses. However, their dyslexic group might have compensated their deficits, especially those who enroll in university education. In dyslexic children, Martinez Perez et al. (2012) reported both item and serial order STM deficits, with these deficits appearing to be independent. The serial order STM deficit was evident when compared to both chronological age and reading age-matched control groups, whereas the item STM deficit was observed solely in relation to the chronological age-matched control group. Staels and Van den Broeck (2014) found that multilingual children diagnosed with dyslexia also exhibited difficulties in both item and serial order STM. However, the authors found that the deficit in the serial-order STM task appeared to be dependent upon the deficit in the item STM task and concluded that serial order STM was not impaired in dyslexia (see also Staels & Van Den Broeck, 2015, who strongly argue against a serial-order impairment in dyslexia). While these questions have mainly been addressed for dyslexia, children with DLD also exhibit verbal STM impairments when performing digit recall tasks and word/non-word recall tasks (Archibald & Gathercole, 2006; Briscoe & Rankin, 2009; Hutchinson et al., 2012; Petruccelli et al., 2012), regardless of the verbal output requirement (Botting et al., 2013).

Overall, there is a consistent impairment of verbal STM in dyslexia and DLD, and these deficits persist until adulthood. The precise causes of this impairment are still under active debate. It appears that for dyslexia, the verbal STM deficits cannot be explained only on the basis of underlying phonological processing impairment given that some studies also show impairment for visuo-spatial STM (Hachmann et al., 2014; Martinez Perez et al., 2015). According to the item/serial-order perspective, dyslexia involves impairment in both item STM, which heavily relies on phonological processing and is specific to verbal

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

material, but also in serial order STM components, which might exhibit impairment across different modalities and materials like music.

Musical STM in dyslexia and DLD

To examine whether the observed STM deficits are specific to verbal material in dyslexia, researchers have explored the potential presence of similar deficits with musical material. Ziegler et al. (2012) demonstrated that dyslexic children exhibit lower performance compared to TD children in recognition STM tasks involving pitch. Furthermore, Atterbury (1985) and Forgeard et al. (2008) reported impaired tonal recognition in dyslexic children. The prevailing explanation for the observed STM impairment in both verbal and musical material in dyslexia suggests that their phonological or pitch representations are degraded (Adlard & Hazan, 1998; Hornickel & Kraus, 2013; Hyde & Peretz, 2004; Ziegler et al., 2009). However, alternative studies have proposed that these representations might actually be intact but more challenging to integrate into memory processes (Boets, 2014; Boets et al., 2013; Ramus & Szenkovits, 2008). To investigate this question, one approach involves simultaneously studying dyslexia and a specific disorder related to music representation/memory: congenital amusia (see section 1.1.3.4.2, page 53-54). Dyslexia and congenital amusia share relevant characteristics such as potential genetic influence and impaired STM processes and/or compromised conscious access to phonological/pitch representations (Couvignou et al., 2023).

Although dyslexia and amusia are estimated to affect approximately 3–7% and 1.5–4% of the general population, respectively (Lindgren et al., 1985; Peretz & Vuvan, 2017), a noteworthy 25–30% of adults diagnosed with dyslexia or amusia also meet criteria for the other disorder (Couvignou et al., 2019). This comorbidity has been confirmed in children aged 7–12 years, where around 34% of dyslexic children were found to have congenital amusia as well

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

(Couvignou & Kolinsky, 2021). These individuals exhibit a combination of traits from both disorders, including slow and error-prone reading, deficient phonological skills (particularly phonological awareness), and struggles in detecting pitch variations in melodies despite having normal hearing, typical nonverbal IQ, and regular music exposure. One hypothesis posits that dyslexia and amusia may represent distinct manifestations of the same underlying STM impairment. Couvignou et al. (2023) addressed this question by testing dyslexic children without amusia, dyslexic children with amusia, and age/reading-level matched controls across three high-level cognitive functions: auditory STM, perceptual awareness, and attention. Interestingly, they found that only auditory serial-order STM was particularly impaired in children with the comorbid dyslexia/amusia phenotype, including in comparison with control children matched for reading level. These findings align with the previously discussed hypothesis that while domain-specific STM deficits may explain impairment in verbal item processing, a broader domain-general serial-order STM impairment could provide a rationale for the observed deficits in both musical and verbal material, as well as the comorbidity of dyslexia and amusia.

Cerebral correlates of auditory STM in dyslexia and DLD

Neuroimaging studies have revealed several structural and functional alterations in dyslexia and DLD. In dyslexia, altered gray matter density has been observed in several regions including the temporal, frontal, and parietal cortices, along with altered white matter density in regions including bilateral temporal and frontal lobes (for a review, see Mascheretti et al., 2017). Numerous functional alterations have been observed for dyslexics' auditory processing of speech sounds at the perceptual level using a variety of tasks. When passively listening to single letters, dyslexics show reduced activation in the superior temporal gyrus compared to TD children (Blau et al., 2010), suggesting altered

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

processing of speech sounds in the ventral auditory stream. Interestingly, increased activation in the superior temporal gyrus is observed when dyslexics perform speech-in-speech recognition compared to TD children, suggesting the need to over-recruit neural resources to overcome speech-in-speech difficulties (Dole et al., 2014). When asked to discriminate vowel length (Steinbrink et al., 2012) or to identify high-pitched non-linguistic stimuli with a speech-like spectrotemporal structure (Gaab et al., 2007), children with dyslexia show reduced activation of the inferior frontal gyrus (IFG) and anterior insula compared to TD children. These results suggest altered fast-temporal auditory processing in dyslexia that affect the perception of acoustic elements characterized by rapid transitions or short durations (e.g., speech sounds). Functional alterations in the prefrontal cortex (PFC) have also been observed in dyslexic individuals while being tested for phonological awareness (the knowledge that speech is composed of syllables and phonemes). Heim et al. (2010) observed reduced activation in the IFG in dyslexic children as compared to TD children while performing phonological decisions. Additionally, Kovelman et al. (2012) observed that 7-13 years old dyslexic children do not recruit the dorsolateral prefrontal cortex (dlPFC) while performing phonological judgements while age-matched TD children and younger children (5-6 years old) do recruit this area. These results suggest a dysfunctional engagement of lateral prefrontal regions in dyslexic children while processing phonological representations.

Structural and functional neuroimaging studies in children with DLD are scarce. Studies investigating structural alterations in individuals with DLD have shown anatomical abnormalities with reports of increased regional volumes in temporal and parietal regions (Soriano-Mas et al., 2009), reduced gray matter in temporal cortices and increased white matter concentration in temporal

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

cortices and in the medial prefrontal lobe (Badcock et al., 2012). The few functional neuroimaging studies performed on children with DLD reported hypo-activation of the posterior superior temporal gyrus while children performed a covert auditory response naming task (Badcock et al., 2012) and covert lexical semantic and phonological tasks (De Guibert et al., 2011). Interestingly, Badcock et al (2012) observed hyperactivation in the insula, extending to the IFG. These studies suggest a common functional alteration of temporal and frontal regions in children with DLD and dyslexia when processing speech material.

Only a few neuroimaging studies investigated auditory STM in children with dyslexia and, to our knowledge, none in children with DLD. Most neuroimaging studies that studied the maintenance of verbal information did so using n-back tasks (see section 1.2.5) that tackles working memory (WM) processes rather than purely STM ones as they entail maintenance and manipulation of information. Beneventi et al. (2010a) used visually presented pictures with which young adolescents with and without dyslexia had to perform a n-back task by using the first or last phoneme of the picture denomination with increasing memory loads. Controls showed increased fMRI activation with memory load in the superior parietal lobule and the IFG while dyslexics did not show a significant increase in activation with increased memory load in these areas. In another study using a n-back task with visually presented letters, Beneventi (2010b) observed that controls and dyslexic children showed increased activity with increasing memory load in the dlPFC and parietal regions but that dyslexics showed a less important increase of activation than controls in the dlPFC. Interestingly, there was increased activity with memory load in the IFG for dyslexics but not for controls, suggesting that in addition with a disrupted allocation of resources in the dlPFC

in dyslexics, they might recruit different frontal regions from the WM network to perform the same task as controls. In line with these findings, Vasic et al (2008) showed that when using a parametric visual-verbal WM task, dyslexics displayed a less important increase of activation with memory load than controls in the dlPFC and a more important increase of activation with load than controls in the IFG. We found only one study that investigated STM specifically by using a Sternberg-like procedure (Beneventi et al., 2009) where participants were asked to report if a target letter was present in a previously visually presented sequence of letters (probe task) or to report if a target sequence of letters was presented in the same order as a previously presented one (serial-order task). In the probe task, dyslexics showed reduced activation in the precentral gyrus compared to controls. In the sequence task, dyslexics showed reduced activation in occipital and parietal regions. Interestingly, when contrasting the serial-order task with the probe task, control participants recruited a wide range of prefrontal regions including the dlPFC, medial PFC, and IFG while dyslexics showed no significant prefrontal engagement. These results obtained with visual-verbal stimuli might reflect the impairment of domain-general serial-order STM observed in dyslexia (Majerus & Cowan, 2016) and amusia (Couvignou et al., 2023).

Certainly due to the noisy nature of fMRI, very few studies have investigated functional alterations related to auditory WM or STM in individuals with dyslexia or DLD. One study investigated auditory WM with fMRI in adults with a history of dyslexia and control participants using pseudowords or tones (Conway et al., 2008). In the pseudowords condition, the task consisted in hearing a pseudoword of 2, 3, 4, or 5 syllables, covertly segment it, maintain the information of the number of syllables and after a delay, hearing another pseudoword and indicate by a button press if the two

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

pseudowords had an equal number of phonemes. In the tone condition, participants performed the same task with tone sequences. For the tone condition, controls showed significant activation in the superior temporal gyrus, the superior frontal gyrus (SFG), and parietal regions, while dyslexics only recruited the superior temporal gyrus. In the pseudowords condition, controls recruited temporal regions, frontal regions (SFG, dlPFC), and parietal regions while dyslexics only recruited the superior temporal gyrus. While these results suggest an absence of recruitment of relevant WM-associated frontal and parietal regions in dyslexics, it is important to note that the task used in the study might have been particularly challenging for dyslexics as an important perceptual awareness component, known to be impaired in dyslexics, was necessary to perform the task (segmentation of pseudowords and tones).

Functional near-infrared spectroscopy (fNIRS) is particularly well-suited to explore auditory STM in children with dyslexia and DLD, as it provides a quiet and less restrictive environment, making it more comfortable and suitable for children and clinical populations who may find fMRI distressing (see section 1.2.3). Sela et al. (2012) compared adults with compensated dyslexia (controlled for phonological awareness) with controls using a visual-verbal n-back task while recording frontal regions with fNIRS. Dyslexics showed reduced overall HbO concentrations in the lateral PFC. Unexpectedly, the authors also observed increased HbR concentrations with increased memory load for both groups. Dyslexics showed lower HbR concentrations than TD children overall. It should be noted that the authors did not filter their raw data for very low frequencies and did not use preprocessing measures to overcome systemic effects, making the interpretation of these results difficult. One recent fNIRS study manipulated memory load with an auditory n-back task in DLD and TD children (Hancock et al., 2023) and found an increase of activity with increasing

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

memory load in the dlPFC and a decrease of activity with increasing memory load in bilateral inferior parietal lobules (IPLs) in the TD group but not in the DLD group. These fNIRS results are overall in line with previously described fMRI studies that show that individual with dyslexia or DLD tend to be disrupted in the recruitment of WM-related regions notably in prefrontal areas.

To our knowledge, no neuroimaging study has investigated auditory STM for musical and verbal material in children with learning disorders. As shown in section 4, the delayed matching to sample task (DMST) is a suitable paradigm to manipulate memory load for STM with musical and verbal material. Moreover, using DMST with musical and verbal material, we were able to show that fNIRS was well-suited to explore lateral prefrontal engagement in auditory STM in healthy adults. More specifically, we showed that only the IFG (but not the dlPFC or the SFG) exhibits parametric activations with memory load and different dynamics were observed for musical (decreased activation with increased memory load) and verbal (increased activation with increased memory load) material. The goal of the present study is to explore the engagement of lateral prefrontal regions in school-aged children with and without language-related learning disorders (dyslexia and DLD) using a child-adapted DMST (also used in section 3) where we manipulate memory load for musical and verbal material. In control children, we expect to replicate results already observed in adults, that is a differential recruitment of the IFG depending on memory load. However, as auditory STM for both materials does not mature until late development (see section 3), we cannot rule out the possibility that children rely differently than adults on lateral prefrontal areas to perform auditory STM tasks. In children with dyslexia and DLD, for verbal STM, we expect a disrupted recruitment of lateral PFC regions either with an absence of parametric variation of activation with memory load, or with a

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

reduced one as compared to controls, or even with a different dynamic parametric activation (e.g., decreasing activation with increasing memory load). For musical STM, we expect that dyslexic/DLD children affected by amusia (estimated at 30% of dyslexic children, based on Couvignou & Kolinsky, 2021 and Couvignou et al., 2023) show disrupted parametric activation of prefrontal regions with memory load for the musical material. Conversely, dyslexic/DLD children without amusia should show the same hypothesized parametric activation of prefrontal regions with memory load as controls for the musical material.

Methods

Participants

For the present study, two groups of children are currently being recruited. For both groups, children are included if they do not present any psychiatric or neurological disorder, any medical treatment that would affect the central nervous system, any musical training of more than one year, or any auditory peripheral disorder, as assessed by a pure-tone audiometry (separately for the two ears, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 6000 Hz, 8000 Hz). Moreover, they are excluded if they present any diagnosed attention-deficit/hyperactivity disorder (ADHD), developmental verbal dyspraxia (DVD), autism spectrum disorder (ASD), or stuttering. Children from 1st Grade's last quarter (French education system acquired reading period, i.e., end of "Cours Préparatoire, CP") to 7th grade (i.e., "5^{ème}" in the French system) could be recruited. For the learning disorder group, children are included if they present a dyslexia and/or developmental language disorder (DLD) diagnosed by a certified professional. Children in the control group have to be exempt of any dyslexia and DLD diagnosis. Both parents or legal guardians had to give their informed consent before testing and filled a questionnaire about the child's

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

level of education, their own level of education, the child's laterality, medical background and diagnosis. All study procedures were approved by a national ethics committee (CPP Ile de France VI, ID RDC 2018-A02670-55). As of today, 11 children were included in the control group (7 girls, one left-handed, mean age = 8.6 years, sd = 1.3 years, min = 7.6 years, max = 11.4 years) and one child was included in the learning disorder group (age = 13 years).

Stimulus construction and task design

Musical and verbal stimuli were the same as in section 3 and 4: six musical tones belonging to the C major scale (C2, E2, G2, B2, D3, F3) and six syllables (/fi/ /gu/ /ly/ /mø/ /te/ /za/).

All tasks were STM tasks. There were two memory loads (ML) that differed in sequence length: ML1 and ML2. Sequence length were shorter than in section 4 to make it feasible for children. For the musical material, the two MLs consisted in respectively three- and four-item sequences; for the verbal material, they consisted in respectively four- and five-item sequences. Within a trial, S1 and S2 always had the same number of items. Item duration, inter-stimulus interval, delay duration, response time-window and jittered inter-trial interval were as described in section 4. Due to the ML manipulation, the durations of S1-delay-S2 were respectively 9400ms and 10600 ms for ML1 and ML2 for the musical material, and respectively 10600 ms and 11800 ms for ML1 and ML2 for the verbal material.

Twelve S1 sequences were created for each ML level and for each material (six for same trials, six for different trials). Since there were no sequences over 6-item length, all items were different within a sequence. When S2 was different, two adjacent items were switched (instead of introducing a new item) thus systematically changing S2 contour for the musical material.

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

Any item could be switched with the next one, except for the first item. There was an equiprobable number of sequences with each position of item switch (e.g., for musical ML2 “different” S2 sequences, there were three trials with a switch between the 2nd and the 3rd item and three trials with a switch between the 3rd and the 4th item). Additionally, sixteen 24400- to 28400-ms randomly jittered silent trials were generated to intersperse within testing blocks (four silent trials per block, see below).

fNIRS data acquisition and montage

The absorption of near-infrared light was measured at 760- and 850-nm wavelengths at a sampling frequency of 7.81 Hz using a continuous-wave NIRScout device (NIRx Medical Technologies, LLC). The data were collected using the NIRStar 15.3 acquisition software. Eight light sources and eleven light detectors were attached to a cap with a 10-20-system marking for probe placement. Additionally, eight 8-mm short-distance channels (one for each source) recorded systemic signal.

We adapted the montage from the one used in adults in section 4. We kept the same eight channels recording bilateral dlPFC and the six channels recording bilateral IFGs (that were centered above the triangular part). We removed the channels recording the medial PFC, as we observed the absence of their involvement in the adult study, and added one channel over each hemisphere that had maximum specificity for the opercular part of the IFG and one channel over each hemisphere that had maximum specificity for the orbital part of the IFG. The final montage thus included 20 measurement channels with distances ranging from 2.9 to 4.1 centimeters (the montage is represented on Figure 1 and see Supplementary Table S1 for maximum recording specificity of each channel). Data from two channels that had the highest specificity for the precentral gyrus were not analyzed as they were not in the scope of the

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

current study. For clarity purposes and since the data presented here are only preliminary, we will present analyses only for the channels already recorded in section 4 (dlPFC and IFG centered over the triangular part).

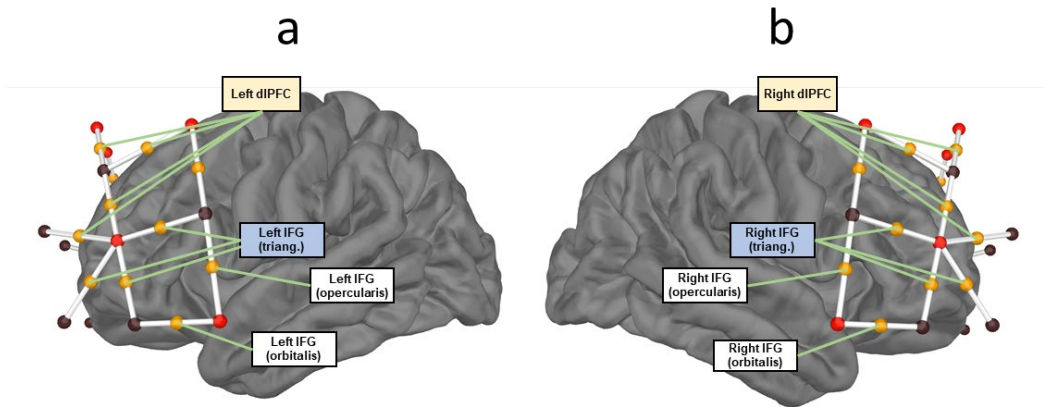


Figure 1 : Location of sources (black spheres) and detectors (red spheres) and their corresponding midpoint channels (white lines with orange spheres) for (a) left and (b) right views of the brain. The 3 channels recording the triangular part of bilateral IFGs (triang.) and the 4 channels recording the dlPFC are connected with green for each hemisphere on the lateral views. For completeness, the channel recording the opercular part (opercularis) of the IFG and the channel recording the orbital (orbitalis) are represented the same way for each hemisphere. Brain and montage figures were generated using the MNE-NIRS python library (Luke et al., 2021).

Procedure

The whole experiment was divided in two sessions. In the first screening session that lasted about one hour, children first underwent two reading tests from the EVALEO battery (Launay et al., 2018), a diagnostic tool used for learning disorders diagnosis by French speech-language therapists. The reading tests aimed at obtaining each child's reading age for later matching between the learning disorder and control groups. Next, children underwent pure-tone audiometry. Finally, children underwent an adapted version of the Montreal Battery of Evaluation of Musical Abilities (MBEMA, Peretz et al., 2013). This test, adapted from the adult's Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al., 2003), aims at diagnosing amusia in children by testing

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

their musical perception and STM abilities on five subscales (scale, contour, interval, rhythm, and memory, overall score calculated as the mean score of the five subscales, maximum score = 20). We adapted the implementation of the MBEMA to use the same story-telling with animals than for the main STM experiment (see section 3 and below). Additionally, children underwent a child-adapted Speech, Spatial and Qualities of Hearing questionnaire (Kid-SSQ; Coudert et al., 2023), a tool of rapid screening of children difficulties in various auditory domains including speech-in-noise perception.

In the second session (on a different day) that lasted around one hour, children were recorded with fNIRS while they performed the STM tasks. The experimenter first measured the participant's head circumference to determine cap size. Cap alignment was verified and adjusted if needed so that the probe at Cz was located halfway in the nasion-to-inion and the tragus-to-tragus measurements. The participant was then led into a dimly lit, sound-attenuated booth where, if needed, the participant's hair was moved around the optode locations using a thin wooden stick to provide clear access to the scalp. Optodes were then placed according to the pre-established montage (Figure 1). fNIRS signal was calibrated, checked for quality, and optode placement on the scalp was readjusted until a satisfying signal quality was obtained before proceeding.

Then, the experimenter gave the task instructions with the same cover story as section 3, corresponding to the visual stimuli displayed on the computer screen during the task (Figure 2a): during the first sequence, a cartoon picturing an elephant teacher would appear on the computer screen and during the second sequence, a cartoon of a nice blue baby-elephant would appear on the left and a cartoon of a grimacing red baby-monkey would appear on the right. After the second sequence, a question mark would appear on the screen between the two cartoons. Children were given the instruction that the baby-elephant was

always repeating correctly the sequence produced by the elephant-professor and that the grimacing-monkey was always repeating incorrectly. They had to give their response with the computer keyboard, placed in front of them at a comfortable distance. They had to press the ‘q’ key with their left hand for ‘same’ response (indicated with a blue sticker, corresponding to the blue baby-elephant) and the ‘m’ key with their right hand for ‘different’ response (indicated with a red sticker, corresponding to the red baby-monkey). For a trial example, see Figure 2a.

Presentation of stimuli and recording of behavioral responses was controlled by Presentation© software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Triggers were sent from Presentation© to the fNIRS acquisition system using a parallel port. Children sat approximately 50 cm from the computer screen, auditory stimuli were played through Z200 Logitech® speakers at a comfortable intensity (55-60 dB SPL A-weighted measured at the location of the participants’ ears with a Brüel & Kjær type 2239 sonometer), each speaker was placed approximately 70 cm from the corresponding ear.

All children underwent four blocks of sixteen trials, each block containing 12 stimulation trials and 4 silent trials on the 1st, 6th, 11th, and 16th position as in section 4. Each block could randomly contain stimulation trials from any of the four conditions: ML1 music, ML2 music, ML1 verbal, ML2 verbal. In any block, there could be no more than three consecutive “same” (or “different”) trials, no more than three consecutive trials of the same memory load, and no more than four consecutive trials of the same material. Each of the six blocks lasted around 5 minutes. The order of test blocks was counterbalanced across participants using a Latin square balanced for first-order carryover effects using the `crossdes` R package (Sailer, 2022). There was

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

a four-trial training block at the beginning of the experiment, with one trial from each ML condition, half of them “same” and half of them “different”. The experimenter stayed with the child in the experiment room during recordings, ensuring the child’s comfortableness and focus on the task.

Behavioral data analysis

Behavioral analyses presented were carried out in the same way as in section 4. Measures of d-prime (d') and criterion (c) were obtained according to Signal Detection Theory (SDT) for each task, material, and participant (Macmillan & Creelman, 2005). Hit corresponded to a correct answer for a different trial. False alarm corresponded to an incorrect answer for a same trial. d' , or sensitivity, was calculated using the *psycho* package (Makowski, 2018) as the z-score of False Alarms subtracted from the z-score of Hits. The criterion was calculated as the mean z-score of Hits and False-alarm rates multiplied by minus one and reflect an observer’s bias to say yes (in our case “different”) or no (“same”), an unbiased observer having a value around 0. A liberal bias (tendency to say “different”) results in a negative c , a conservative one results in positive c . Correction of extreme values was made following Hautus (1995) who recommends the use of a log-linear rule that consists of increasing each cell frequency of the contingency table by 0.5, irrespective of the content of each cell. Furthermore, we analyzed the response times (RT) of participants after the end of S2 for correct trials only, averaged separately for each task, material, and each type of trial (same/different).

Analyses were conducted using Bayesian Statistics (Wagenmakers et al., 2018). We report Bayes Factor (BF_{10}) as a relative measure of evidence of an effect compared to the null model. Traditionally, a BF_{10} between 1 and 3 is considered as weak evidence for the tested model, between 3 and 10 as positive evidence, between 10 and 100 as strong evidence and higher than 100 as decisive

evidence. Similarly, to interpret the strength of evidence in favor of the null model, a BF_{10} between 0.33 and 1 is considered as weak evidence, a BF between 0.01 and 0.33 as positive evidence, a BF between 0.001 and 0.01 as strong evidence and a BF lower than 0.001 as decisive evidence (Lee & Wagenmakers, 2014). For clarity purposes, we report information about the best model only.

Using the R *BayesFactor* package (Morey & Rouder, 2022), d' and c were submitted to a Bayesian repeated-measure ANOVA including memory load (ML1/ML2), material (two levels: musical and verbal), and their interaction as fixed factors. Overall, four models were tested (memory load, material, memory + material, memory + material + memory load:material) and compared to the null model. As recommended by Van Den Bergh et al. (2022), participants were added to all models as random factors using the *lmBF* function of the *BayesFactor* package. Paired Bayesian t-tests were performed as post-hoc tests if the best model included the interaction. Correct RTs were submitted to the same Bayesian ANOVA, with the addition of the type of trial (same/different) factor.

Additionally, one-sample Bayesian t-tests against 0 were performed on the criterion for each task and material.

Furthermore, we report the results of the analysis of effects using the *bayesfactor_inclusion* function from the R *bayestestR* package (Makowski et al., 2019) that compares models that incorporate a specific effect, such as a factor or an interaction, with equivalent models without the given effect. The resulting measure, $BF_{inclusion}$, serves as a relative indicator of the evidence favoring the inclusion of a factor.

Analysis described in the present section were carried out on the 11 control children while data from the dyslexic child will be described numerically.

fNIRS data pre-processing and deconvolution

The same procedure was used as in section 4. fNIRS data were pre-processed using the NIRS Brain-AnalyzIR toolbox (Santosa et al., 2018) and custom-written scripts. First, data were trimmed with a custom-written script to remove signal parts for which the participant did not perform the task (instructions, breaks between blocks etc.). Then raw intensity signals were converted to changes in optical density. To correct for motion artifacts from excessive head movements, we applied Temporal Derivative Distribution Repair (TDDR, Fishburn et al., 2019). Corrected optical density were then band-pass filtered between 0.01 and 0.2 Hz to remove cardiac (~ 1.2 Hz) and respiratory activity (~ 0.25 Hz). Finally, corrected and filtered optical densities were transformed into (de)oxygenated hemoglobin concentrations using the modified Beer-Lambert Law.

Data were then processed with a General Linear Model (GLM) using a Finite Impulse Response (FIR) model to deconvolute the signals from successive trials for each channel. To do so, 29 one-second boxcar regressors were fitted around delay onsets to encompass the total duration of stimulation (-8 seconds to 20 seconds around delay onset) for each task and material. A boxcar regressor per block (encompassing the entire twelve stimulation trials and four silence trials) was added to account for possible HbO/HbR signal changes across blocks of recordings. Finally, data from all short-channels (eight HbO and eight HbR measures) were orthogonalized and added as regressors of no-interest in the GLM in order to further clean the signal from systemic components (Luke et al., 2021). Overall, for each chromophore (HbO/HbR), each recording channel of each participant was regressed using an Ordinary Least Squares (OLS) GLM with a design matrix including block regressors, 29 1-second boxcar regressors for each task and material (thus 116 regressors for the FIR models), and all

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

short-channel signal (16 regressors). Note that no baseline correction is needed in the FIR GLM approach as silent trials, for which no regressors are fitted in the GLM, act as implicit baseline (Cairo et al., 2004; Kharitonova et al., 2015).

fNIRS data analyses

Deconvoluted data were then analyzed using a Bayesian ANOVA on beta coefficients for each 1-s time window of the FIR models and each ROI. Four ROIs (see Figure 1 and supplemental Table S1) were created with the channels showing the largest specificity for the left or right triangular part of the IFG (3 channels in each hemisphere) and the left or right dlPFC (4 channels in each hemisphere). For each participant, time point, memory load, and material, betas were averaged across channels making up each ROI. These ROI-averaged betas were then tested across participants, for each time point from -8 to 12 seconds around delay onset (i.e., until 6 seconds after the end of S2, as the deconvoluted signal would then include the motor response which we did not intend to analyze) for each ROI. For each timepoint, two paired Bayesian t-tests were performed on average beta between the first level of memory load (ML1) and the second (ML2), for each material separately. We did not perform Bayesian ANOVA with the memory load and the material in the same model as in section 4 because of the small number of participants in these preliminary data. For the same reason, analysis of fNIRS data were carried out without dividing children into age groups.

We report in the main text HbO and HbR results for the 11 control children.

Results

Screening tests

During the pure-tone audiometry, no child presented a hearing threshold higher than 30 dB at any tested frequency (250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 6000 Hz, 8000 Hz) in both ears.

For the MBEMA, all children presented an MBEMA higher than the normative cut-off score from Peretz et al. (2013), see Figure 2b for individual results. However, one control child displayed a MBEMA score at cut-off.

Children's reading level was evaluated with the EVALEO battery (Launay et al., 2018) that provides, for two reading comprehension tasks (reading of a meaningless text and a text with meaning), a score for which children show "normal" performance when they score between the 21st to 80th percentile of a normative population of their corresponding age and education level. All children tested so far, including the dyslexic child, performed for both reading tests within "normal" scores.

The Kid-SSQ questionnaire scores are yet to be analyzed.

STM: Behavioral results

For d' in control participants (Figure 2c), the best model explaining the data included both fixed factors (memory load and material, strong evidence, $BF_{10} = 94.32$) but not the interaction between both. The analysis of effects across matched models revealed strong evidence for the memory load ($BF_{inclusion} = 40.43$) factor and positive evidence for the material ($BF_{inclusion} = 6.6$) factor. d' was higher for ML1 as compared to ML2 and higher for the verbal material as compared to the musical material.

For correct RTs in control participants (Figure 2d), the best model was the null model, no evidence for an effect of the material, memory load, type of trial or their interaction was found (all $BF_{inclusion} < .104$).

For the criterion (c), the best model was the null model, no evidence for an effect of the material, memory load or their interaction was found (all $BF_{inclusion} < .34$). Bayesian t-tests against 0 revealed weak evidence for a positive bias for the musical material in the ML1 condition ($BF_{10} = 1.08$). All other Bayesian t-tests revealed weak evidence for the null model (all $BF_{10} < 0.33$).

The only dyslexic child (aged 13, thus older than the control participants) included in the study so far exhibited a higher d' than the averaged d' of controls for both ML1 and ML2 conditions in the musical material and only for the ML1 condition in the verbal material. However, she showed a lower d' than the average of controls in the ML2 verbal condition (see Figure 2c), where her performance was in the range of the youngest children. The response times (Figure 2d) and criterion obtained by this child did not seem to differ from controls.

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

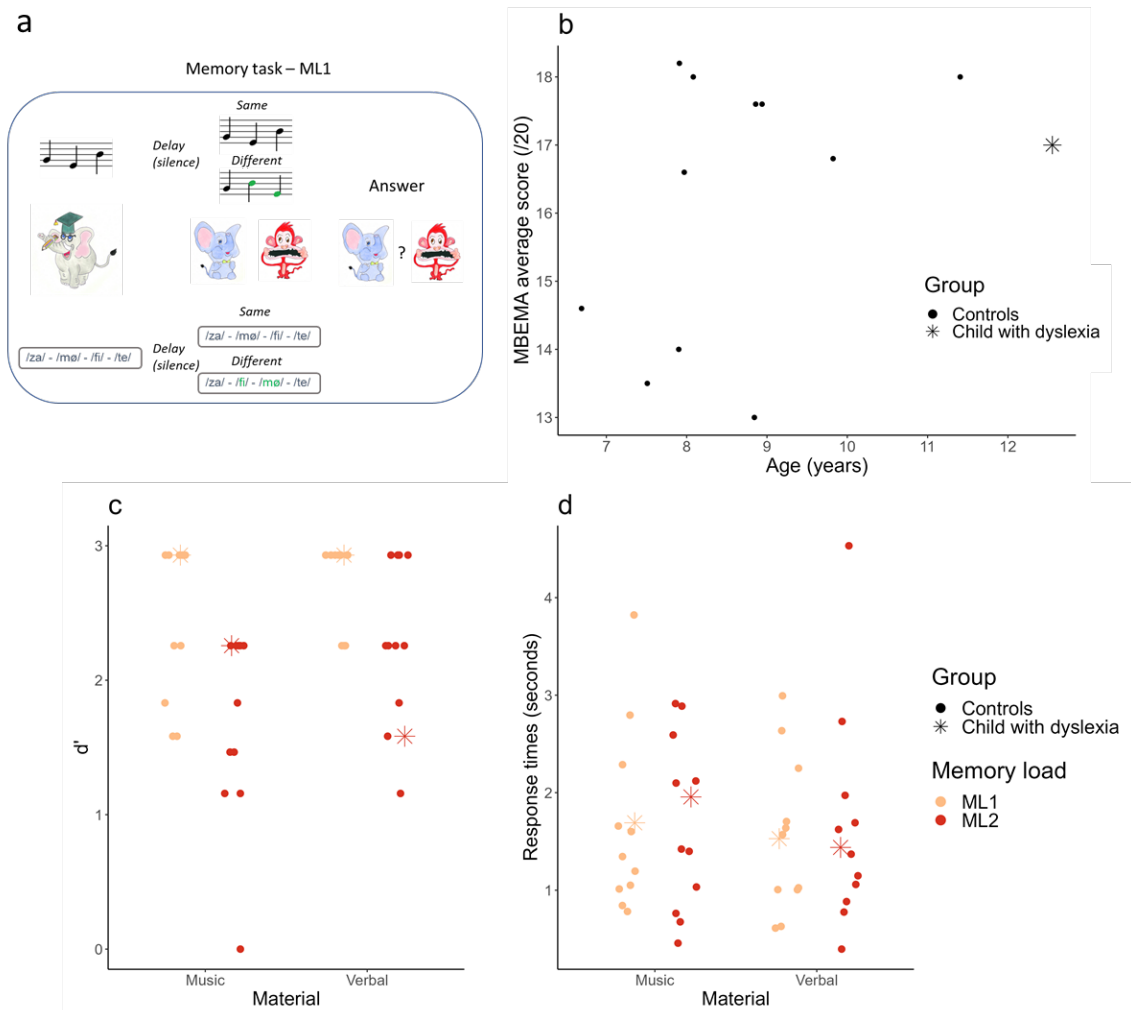


Figure 2: Tasks and average performance as a function of memory load (ML1, orange color/ML2, red color), material (musical/verbal), and group (controls, filled circle, $n = 11$, aged 7.6-11.4 years; learning disorder, star-shaped point, $n=1$, aged 13 years).

The results of the child included in the learning disorder group is displayed for illustration purposes only and was not included in any analysis. (a) Example of a memory task trial for the ML1 condition for both materials along with the visual stimuli that were displayed during the task as cover-story. (b) individual MBEMA averaged scores (over five subscales, maximum score = 20) as a function of age. (c) Individual sensitivity scores (d') as a function of material and memory load, for the control children and the dyslexic child. (d) Individual averaged response times for correct trials (time in millisecond that children spent after the end of S2 before giving a “same” or “different” answer) as a function of material and memory load, for the control children and the dyslexic child.

STM: fNIRS results

Results of the HbO and HbR deconvoluted signal within the targeted ROIs are presented in the next section (Figure 3 and Figure 4). As a reminder,

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

we performed Bayesian t-tests between memory load conditions separately for each material.

HbO results



Figure 3 : (a) and (c) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbO data for all children from the control group ($n=11$), across four ROIs (left and right triangular IFG, left and right dlPFC), in a time window ranging from -8 to 20 seconds around delay onset (grey dotted vertical line), for the two memory load levels (ML1, orange color/ML2, red color) for the musical material (a) and verbal material (c). For clarity purposes, S1, delay, and S2 durations are indicated with double-headed arrows. (b) and (d): time-course representation of the results of Bayesian t-tests between the two memory load levels for each material. The dark pink shade indicates evidence for a difference between memory loads ($BF_{10} > 1$), light pink shade indicate evidence for the null model ($BF_{10} < 1$) for the musical (b) and verbal (d) material. No statistics were performed beyond 12 seconds after delay onset (end of S2 + peak of the hemodynamic response, ~ 6s). D: silent retention delay; ML: memory load.

For the musical material, all comparisons showed weak evidence for the null model ($.3 < BF_{10} < .84$) and thus no evidence for a difference between ML1 and ML2 was found. By visually inspecting the HbO results (Figure 3), a parametric decrease of averaged betas seems to take place in the left IFG (after the delay) and the left dlPFC (overall) with lower betas for the ML1 as compared to the ML2 condition. However, a small parametric increase of betas with memory load seem to take place in the right dlPFC with higher betas for the ML2 condition than for the ML1 condition.

For the verbal material, all comparisons in bilateral IFG showed weak evidence for the null model ($.3 < BF_{10} < .75$) and thus no evidence for a difference between ML1 and ML2 was found. However, a visual inspection seems to reveal a parametric increase with memory load of the averaged beta in left and right IFGs 3 to 10 seconds after delay onset (corresponding to the hemodynamic response to the delay, starting at 0 and finishing at 6 seconds) with higher betas for the ML2 conditions than the ML1 condition. For the left dlPFC, we found weak to strong evidence -8 to -1 seconds and 6 to 9 seconds around delay onset for a difference between ML1 and ML2 ($1.05 < BF_{10} < 10.07$) with higher betas for ML2 than for ML1. For the right dlPFC, we found weak evidence 6 to 8 seconds after delay onset for a difference between ML1 and ML2 ($1.72 < BF_{10} < 2.22$) with higher betas for ML2 than for ML1.

HbR results

For the musical material, comparisons in the left IFG showed weak evidence for the null model ($.3 < BF_{10} < .67$) thus no evidence for a difference between ML1 and ML2 was found. By visually inspecting results, an apparent parametric decrease with memory load seem to take place in the left IFG with lower betas for the ML2 condition compared to the ML1 condition. In the right IFG, for the musical material, we found weak evidence for a difference between

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

ML1 and ML2 ($BF_{10} = 1.12$) 8 seconds before delay onset, with lower betas for ML2 than for ML1. Comparisons in the left dlPFC showed weak evidence for the null model ($.3 < BF_{10} < .44$) thus no evidence for a difference between ML1 and ML2 was found. In the right dlPFC, for the musical material, weak evidence was found -8 and -7 seconds before delay onset for a difference between ML1 and ML2 ($1.01 < BF_{10} < 1.08$) with lower betas for ML2 than for ML1.

For the verbal material, in the left IFG we found weak evidence for a difference between ML1 and ML2 ($BF_{10} = 1.29$) 10 seconds after delay onset with lower betas for ML2 than for ML1. We can observe numerically a decrease of averaged betas with increasing memory load after delay onset in the left IFG with lower betas for the ML2 condition compared to the ML1 condition. Comparisons in the right IFG showed weak evidence for the null model ($.3 < BF_{10} < .42$) thus no evidence for a difference between ML1 and ML2 was found. In the left dlPFC, weak to positive evidence was found -2 to -1 seconds and 8 to 12 seconds around delay onset for a difference between ML1 and ML2 ($1.24 < BF_{10} < 3.93$) with lower betas for ML2 than for ML1. Comparisons in the right dlPFC showed weak evidence for the null model ($.3 < BF_{10} < .75$) thus no evidence for a difference between ML1 and ML2 was found.

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.



Figure 4 : (a) and (c) Average betas (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbR data for all children from the control group ($n=11$), across four ROIs (left and right triangular IFG, left and right dIPFC), in a time window ranging from -8 to 20 seconds around delay onset (grey dotted vertical line), for the two memory load levels (ML1/ML2) for the musical material (a) and verbal material (c). For clarity purposes, S1, delay, and S2 durations are indicated with double-headed arrows. (b) and (d): time-course representations of the results of Bayesian t -tests between the two memory load levels for each material. The dark pink shade indicate evidence for a difference between memory loads ($BF_{10} > 1$), light pink shade indicate evidence for the null model ($BF_{10} < 1$) for the musical (b) and verbal (d) material. No statistics were performed beyond 12 seconds after delay onset (end of S2 + peak of the hemodynamic response, $\sim 6s$). D: silent retention delay; ML: memory load.

Discussion

In the current section, we presented preliminary behavioral and fNIRS results from 11 control children and behavioral results for one child with a

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

diagnostic of dyslexia, from an ongoing study aiming at characterizing the potential functional alterations of auditory STM for musical and verbal material in lateral PFC regions in children with dyslexia and DLD. Behavioral data showed that control children were able to perform the task, as expected from section 3. As also expected, we observed a parametric decrease of performance for increasing memory load. In contrast with the adult study (section 4), we observed lower performance for the musical material than for the verbal material. In the attempt to equalize performance between materials, we reduced the sequence-length difference between equivalent memory loads for verbal and musical material (i.e., in the adult study, for music, 4, 5 and 6 items for ML1/ML2/ML3 respectively and for verbal 6, 7, and 8 items; in the present study, for music 3 and 4 items for ML1/ML2 respectively and for verbal 4 and 5 items). Ultimately, we did not reach equal performance between materials and memory loads but it seems that we reduced the material effect compared to the adult study. Additionally, there was no parametric effect in response times, in contrast with the adult study. This is probably due to the fact that adults were given the instruction that they had 3 seconds to give their response while children had an unlimited amount of time to give their response with no specific instruction thus less pressure to produce speeded responses. Even though no conclusion can be drawn from a single subject, it is interesting to note that the only dyslexic child included in the study so far, who is not amusic, seem to show a greater memory load effect for the verbal material than for the musical material that was not observable in control children.

The ROI analysis of fNIRS HbO data in controls revealed evidence for a parametric effect of the memory load in bilateral dlPFC for the verbal material during the retention period, with higher activation for higher memory load. Additionally, tendencies for a parametric effect for the verbal material

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

can be observed in the IFG. A tendency of an inverse parametric effect can be observed in the right IFG for the musical material with higher activation for the lowest memory load, as it has been observed in the adult's study. HbR data also show a tendency for larger decrease (reflecting higher activation) for the highest memory load in the left dlPFC and IFG. These results suggest that children tend to recruit a broader lateral PFC network involving both the IFG and the dlPFC when memory load increases in auditory STM, in contrast with adults who only recruited the IFG. These preliminary results must be taken with great caution as indicated by the weak evidence levels of the Bayesian analysis. Indeed, the analysis only concerns 11 children of a somewhat large age range (7.5 to 11.4 years) but are promising for the next steps.

The next steps include recruiting up to 20 children with learning disorders and 30 control children with the aim to have enough data from control children to match them in chronological- or reading-age with children with learning disorders for later analysis. We first intend to compare fNIRS data in the control group across three age groups (e.g. 6-7, 8-9, 10-12 years old) in order to verify potential developmental effects on the hemodynamics of children. As maturation of auditory STM occurs (see section 3), we cannot exclude the possibility of changes in the recruitment of prefrontal regions in auditory STM across development. Additionally, vascular maturation in children's brain might provoke hemodynamic changes between younger and older children. Based on these observations, we will compare fNIRS data between the controls group and the learning disorder group, either by dividing them into age groups or all together. Moreover, we will compare them based on data acquired during the screening session. We will analyze data separately for children presenting learning disorder and amusia and for children presenting learning disorder without amusia. Moreover, we will be able to explore the potential impact of

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

listening difficulties on the engagement of prefrontal regions in auditory STM by computing correlations between children's score at the SSQ questionnaire and the amplitude of the memory load parametric effect.

References

- Adlard, A., & Hazan, V. (1998). Speech Perception in Children with Specific Reading Difficulties (Dyslexia). *The Quarterly Journal of Experimental Psychology Section A*, *51*(1), 153-177. <https://doi.org/10.1080/713757570>
- Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675-693.
- Atterbury, B. W. (1985). Musical Differences in Learning-Disabled and Normal-Achieving Readers, Aged Seven, Eight and Nine. *Psychology of Music*, *13*(2), 114-123. <https://doi.org/10.1177/0305735685132005>
- Badcock, N. A., Bishop, D. V. M., Hardiman, M. J., Barry, J. G., & Watkins, K. E. (2012). Co-localisation of abnormal brain structure and function in specific language impairment. *Brain and Language*, *120*(3), 310-320. <https://doi.org/10.1016/j.bandl.2011.10.006>
- Beneventi, H., Tønnessen, F. E., & Ersland, L. (2009). Dyslexic Children Show Short-Term Memory Deficits in Phonological Storage and Serial Rehearsal : An fMRI Study. *International Journal of Neuroscience*, *119*(11), 2017-2043. <https://doi.org/10.1080/00207450903139671>
- Beneventi, H., Tønnessen, F. E., Ersland, L., & Hugdahl, K. (2010a). Executive working memory processes in dyslexia : Behavioral and fMRI evidence: Working memory deficit in dyslexia. *Scandinavian Journal of Psychology*, *51*(3), 192-202. <https://doi.org/10.1111/j.1467-9450.2010.00808.x>
- Beneventi, H., Tønnessen, F. E., Ersland, L., & Hugdahl, K. (2010b). Working Memory Deficit in Dyslexia : Behavioral and fMRI Evidence. *International Journal of Neuroscience*, *120*(1), 51-59. <https://doi.org/10.3109/00207450903275129>
- Blau, V., Reithler, J., Van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure : A functional magnetic resonance imaging study of dyslexic children. *Brain*, *133*(3), 868-879. <https://doi.org/10.1093/brain/awp308>
- Boets, B. (2014). Dyslexia : Reconciling controversies within an integrative developmental perspective. *Trends in Cognitive Sciences*, *18*(10), 501-503. <https://doi.org/10.1016/j.tics.2014.06.003>
- Boets, B., Op De Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). Intact But Less Accessible Phonetic Representations in Adults with Dyslexia. *Science*, *342*(6163), 1251-1254. <https://doi.org/10.1126/science.1244333>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

- Botting, N., Psarou, P., Caplin, T., & Nevin, L. (2013). Short-Term Memory Skills in Children With Specific Language Impairment : The Effect of Verbal and Nonverbal Task Content. *Topics in Language Disorders, 33*(4), 313-327. <https://doi.org/10.1097/01.TLD.0000437940.01237.51>
- Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological Processing, Language, and Literacy : A Comparison of Children with Mild-to-moderate Sensorineural Hearing Loss and Those with Specific Language Impairment. *Journal of Child Psychology and Psychiatry, 42*(3), 329-340. <https://doi.org/10.1111/1469-7610.00726>
- Briscoe, J., & Rankin, P. M. (2009). Exploration of a 'double-jeopardy' hypothesis within working memory profiles for children with specific language impairment. *International Journal of Language & Communication Disorders, 44*(2), 236-250. <https://doi.org/10.1080/13682820802028760>
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order : A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551-581. <https://doi.org/10.1037/0033-295X.106.3.551>
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*(4), 627-652. <https://doi.org/10.1016/j.jml.2006.08.005>
- Cairo, T. A., Liddle, P. F., Woodward, T. S., & Ngan, E. T. C. (2004). The influence of working memory load on phase specific patterns of cortical activity. *Cognitive Brain Research, 21*(3), 377-387. <https://doi.org/10.1016/j.cogbrainres.2004.06.014>
- Conway, T., Heilman, K. M., Gopinath, K., Peck, K., Bauer, R., Briggs, R. W., Torgesen, J. K., & Crosson, B. (2008). Neural substrates related to auditory working memory comparisons in dyslexia : An fMRI study. *Journal of the International Neuropsychological Society, 14*(4), 629-639. <https://doi.org/10.1017/S1355617708080867>
- Coudert, A., Reilly, K. T., Truy, E., Moulin, A., & Gaveau, V. (2023). A new tool for the assessment of speech understanding and spatial hearing difficulties in children : The Kid-SSQ questionnaire. *European Archives of Oto-Rhino-Laryngology, 280*(8), 3557-3566. <https://doi.org/10.1007/s00405-023-07846-9>
- Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia, 155*, 107811. <https://doi.org/10.1016/j.neuropsychologia.2021.107811>
- Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology, 36*(1-2), 1-17. <https://doi.org/10.1080/02643294.2019.1578205>
- Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology, 1-47*. <https://doi.org/10.1080/09297049.2022.2162031>
- De Guibert, C., Maumet, C., Jannin, P., Ferré, J.-C., Tréguier, C., Barillot, C., Le Rumeur, E., Allaire, C., & Biraben, A. (2011). Abnormal functional

- lateralization and activity of language brain areas in typical specific language impairment (developmental dysphasia). *Brain*, *134*(10), 3044-3058.
<https://doi.org/10.1093/brain/awr141>
- Dole, M., Meunier, F., & Hoen, M. (2014). Functional correlates of the speech-in-noise perception impairment in dyslexia: An MRI study. *Neuropsychologia*, *60*, 103-114. <https://doi.org/10.1016/j.neuropsychologia.2014.05.016>
- Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS. *NeuroImage*, *184*, 171-179.
<https://doi.org/10.1016/j.neuroimage.2018.09.025>
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 383-390. <https://doi.org/10.1525/mp.2008.25.4.383>
- Gaab, N., Gabrieli, J., Deutsch, G., Tallal, P., & Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: An fMRI study. *Restorative neurology and neuroscience*, *25*(3-4), 295-310.
- Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259-272.
<https://doi.org/10.1007/BF03174081>
- Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, *7*(4), 821-841. <https://doi.org/10.1007/s12187-013-9230-6>
- Hachmann, W. M., Bogaerts, L., Szmalec, A., Woumans, E., Duyck, W., & Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of Dyslexia*, *64*(2), 121-136.
<https://doi.org/10.1007/s11881-013-0089-5>
- Hachmann, W. M., Cashdollar, N., Postiglione, F., & Job, R. (2020). The relationship of domain-general serial order memory and reading ability in school children with and without dyslexia. *Journal of Experimental Child Psychology*, *193*, 104789. <https://doi.org/10.1016/j.jecp.2019.104789>
- Heim, S., Grande, M., Pape-Neumann, J., Van Ermingen, M., Meffert, E., Grabowska, A., Huber, W., & Amunts, K. (2010). Interaction of phonological awareness and 'magnocellular' processing during normal and dyslexic reading: Behavioural and fMRI investigations. *Dyslexia*, *16*(3), 258-282.
<https://doi.org/10.1002/dys.409>
- Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73-137.
<https://doi.org/10.1006/cogp.1998.0685>
- Hornickel, J., & Kraus, N. (2013). Unstable Representation of Sound: A Biological Marker of Dyslexia. *The Journal of Neuroscience*, *33*(8), 3500-3504.
<https://doi.org/10.1523/JNEUROSCI.4205-12.2013>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

- Hutchinson, E., Bavin, E., Efron, D., & Sciberras, E. (2012). A comparison of working memory profiles in school-aged children with Specific Language Impairment, Attention Deficit/Hyperactivity Disorder, Comorbid SLI and ADHD and their typically developing peers. *Child Neuropsychology*, *18*(2), 190-207. <https://doi.org/10.1080/09297049.2011.601288>
- Hyde, K. L., & Peretz, I. (2004). Brains That Are out of Tune but in Time. *Psychological Science*, *15*(5), 356-360. <https://doi.org/10.1111/j.0956-7976.2004.00683.x>
- Jeffries, S., & Everatt, J. (2004). Working memory : Its role in dyslexia and other specific learning difficulties. *Dyslexia*, *10*(3), 196-214. <https://doi.org/10.1002/dys.278>
- Kharitonova, M., Winter, W., & Sheridan, M. A. (2015). As Working Memory Grows : A Developmental Account of Neural Bases of Working Memory Capacity in 5- to 8-Year Old Children and Adults. *Journal of Cognitive Neuroscience*, *27*(9), 1775-1788. https://doi.org/10.1162/jocn_a_00824
- Kovelman, I., Norton, E. S., Christodoulou, J. A., Gaab, N., Lieberman, D. A., Triantafyllou, C., Wolf, M., Whitfield-Gabrieli, S., & Gabrieli, J. D. E. (2012). Brain Basis of Phonological Awareness for Spoken Language in Children and Its Disruption in Dyslexia. *Cerebral Cortex*, *22*(4), 754-764. <https://doi.org/10.1093/cercor/bhr094>
- Launay, L., Maeder, C., Roustit, J., & Touzin, M. (2018). Evaleo 6-15 : Batterie d'évaluation du langage oral et écrit chez les sujets de 6 15 ans. *Rééducation orthophonique*, *55*, 273. <https://www.orthoedition.com/>
- Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development : Evidence from a longitudinal study. *Developmental Psychology*, *46*(2), 417-427. <https://doi.org/10.1037/a0018540>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling : A practical course*. Cambridge university press.
- Lindgren, S. D., De Renzi, E., & Richman, L. C. (1985). Cross-National Comparisons of Developmental Dyslexia in Italy and the United States. *Child Development*, *56*(6), 1404. <https://doi.org/10.2307/1130460>
- Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, *8*(02). <https://doi.org/10.1117/1.NPh.8.2.025008>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : A user's guide*, 2nd ed. (p. xix, 492). Lawrence Erlbaum Associates Publishers.
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia : The Importance of Serial Order. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.01522>
- Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-Modal Decoding of Neural Patterns Associated with Working Memory : Evidence for Attention-Based Accounts of Working Memory. *Cerebral Cortex*, *26*(1), 166-179. <https://doi.org/10.1093/cercor/bhu189>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

- Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory : The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*(2), 95-119. <https://doi.org/10.1016/j.jecp.2005.07.005>
- Makowski, D. (2018). The psycho Package : An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. <https://doi.org/10.21105/joss.00470>
- Makowski, D., Ben-Shachar, M., & Lüdtke, D. (2019). bayestestR : Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. <https://doi.org/10.21105/joss.01541>
- Martinez Perez, T., Majerus, S., & Poncelet, M. (2012). The contribution of short-term memory for serial order to early reading acquisition : Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *111*(4), 708-723. <https://doi.org/10.1016/j.jecp.2011.11.007>
- Martinez Perez, T., Poncelet, M., Salmon, E., & Majerus, S. (2015). Functional Alterations in Order Short-Term Memory Networks in Adults With Dyslexia. *Developmental Neuropsychology*, *40*(7-8), 407-429. <https://doi.org/10.1080/87565641.2016.1153098>
- Martinez Perez, T., Steve, M., & Martine, P. (2013). Impaired short-term memory for order in adults with dyslexia. *Research in Developmental Disabilities*, *34*(7), 2211-2223. <https://doi.org/10.1016/j.ridd.2013.04.005>
- Mascheretti, S., De Luca, A., Trezzi, V., Peruzzo, D., Nordio, A., Marino, C., & Arrigoni, F. (2017). Neurogenetics of developmental dyslexia : From genes to behavior through brain neuroimaging and cognitive and sensorial mechanisms. *Translational Psychiatry*, *7*(1), e987-e987. <https://doi.org/10.1038/tp.2016.240>
- McGregor, K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981-992. https://doi.org/10.1044/2020_LSHSS-20-00003
- Menghini, D., Finzi, A., Carlesimo, G. A., & Vicari, S. (2011). Working Memory Impairment in Children With Developmental Dyslexia : Is it Just a Phonological Deficity? *Developmental Neuropsychology*, *36*(2), 199-213. <https://doi.org/10.1080/87565641.2010.549868>
- Morey, R. D., & Rouder, J. N. (2022). *BayesFactor : Computation of Bayes Factors for Common Designs*. <https://CRAN.R-project.org/package=BayesFactor>
- Nelson, H. E., & Warrington, E. K. (1980). An investigation of memory functions in dyslexic children. *British Journal of Psychology*, *71*(4), 487-503. <https://doi.org/10.1111/j.2044-8295.1980.tb01762.x>
- Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading Disabilities in SLI and Dyslexia Result From Distinct Phonological Impairments. *Developmental Neuropsychology*, *34*(3), 296-311. <https://doi.org/10.1080/87565640902801841>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder : Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247-1257. <https://doi.org/10.1111/jcpp.12573>
- Ordonez Magro, L., Majerus, S., Attout, L., Poncelet, M., Smalle, E. H. M., & Szmalec, A. (2020). The contribution of serial order short-term memory and long-term learning to reading acquisition : A longitudinal study. *Developmental Psychology*, *56*(9), 1671-1683. <https://doi.org/10.1037/dev0001043>
- Page, M. P. A., & Norris, D. (1998). The primacy model : A new model of immediate serial recall. *Psychological Review*, *105*(4), 761-781. <https://doi.org/10.1037/0033-295X.105.4.761-781>
- Peretz, I. (2013). The biological foundations of music : Insights from congenital amusia. *The psychology of music*, 551-564.
- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders : The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.
- Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*. <https://doi.org/10.3389/fnsys.2013.00030>
- Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*, *25*(5), 625-630. <https://doi.org/10.1038/ejhg.2017.15>
- Petrucelli, N., Bavin, E. L., & Bretherton, L. (2012). Children With Specific Language Impairment and Resolved Late Talkers : Working Memory Profiles at 5 Years. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1690-1703. [https://doi.org/10.1044/1092-4388\(2012/11-0288\)](https://doi.org/10.1044/1092-4388(2012/11-0288))
- Ramus, F. (2003). Developmental dyslexia : Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212-218. [https://doi.org/10.1016/S0959-4388\(03\)00035-7](https://doi.org/10.1016/S0959-4388(03)00035-7)
- Ramus, F., & Szenkovits, G. (2008). What Phonological Deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129-141. <https://doi.org/10.1080/17470210701508822>
- Sailer, M. O. (2022). *crossdes : Construction of Crossover Designs*. <https://CRAN.R-project.org/package=crossdes>
- Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms*, *11*(5), 73.
- Schuchardt, K., Bockmann, A.-K., Bornemann, G., & Maehler, C. (2013). Working Memory Functioning in Children With Learning Disorders and Specific Language Impairment. *Topics in Language Disorders*, *33*(4), 298-312. <https://doi.org/10.1097/01.TLD.0000437943.41140.36>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

- Soriano-Mas, C., Pujol, J., Ortiz, H., Deus, J., López-Sala, A., & Sans, A. (2009). Age-related brain structural alterations in children with specific language impairment. *Human Brain Mapping, 30*(5), 1626-1636. <https://doi.org/10.1002/hbm.20620>
- Staels, E., & Van Den Broeck, W. (2014). Order short-term memory is not impaired in dyslexia and does not affect orthographic learning. *Frontiers in Human Neuroscience, 8*. <https://doi.org/10.3389/fnhum.2014.00732>
- Staels, E., & Van Den Broeck, W. (2015). No solid empirical evidence for the solid (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 650-669. <https://doi.org/10.1037/xlm0000054>
- Steinbrink, C., Groth, K., Lachmann, T., & Riecker, A. (2012). Neural correlates of temporal auditory processing in developmental dyslexia during German vowel length discrimination : An fMRI study. *Brain and Language, 121*(1), 1-11. <https://doi.org/10.1016/j.bandl.2011.12.003>
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245-1260. <https://doi.org/10.1044/jslhr.4006.1245>
- Van Den Bergh, D., Wagenmakers, E.-J., & Aust, F. (2022). *Bayesian Repeated-Measures ANOVA : An Updated Methodology Implemented in JASP* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/fb8zn>
- Vasic, N., Lohr, C., Steinbrink, C., Martin, C., & Wolf, R. C. (2008). Neural correlates of working memory performance in adolescents and young adults with dyslexia. *Neuropsychologia, 46*(2), 640-648. <https://doi.org/10.1016/j.neuropsychologia.2007.09.002>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35-57. <https://doi.org/10.3758/s13423-017-1343-3>
- Wang, X., Xuan, Y., & Jarrold, C. (2016). Using a Process Dissociation Approach to Assess Verbal Short-Term Memory for Item and Order Information in a Sample of Individuals with a Self-Reported Diagnosis of Dyslexia. *Frontiers in Psychology, 7*. <https://doi.org/10.3389/fpsyg.2016.00208>
- World Health Organization. (2015). *International statistical classification of diseases and related health problems* (10th revision, Fifth edition, 2016). World Health Organization.
- Ziegler, J. C., Pech-Georgel, C., George, F., & Foxton, J. M. (2012). Global and local pitch perception in children with developmental dyslexia. *Brain and Language, 120*(3), 265-270. <https://doi.org/10.1016/j.bandl.2011.12.002>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science, 12*(5), 732-745. <https://doi.org/10.1111/j.1467-7687.2009.00817.x>

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD) : A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, *8*(1), 3341. <https://doi.org/10.1038/s41598-018-21716-z>

Supplementary material

Hemisphere	Source	Detector	Cortical structure	Specificity (%)
Left	AF7	F5	IFG (p. Triangularis)	54.20
			MFG	25.02
	F7	F5	IFG (p. Triangularis)	82.28
			IFG (p. Orbitalis)	11.66
	FC5	F5	IFG (p. Triangularis)	68.70
			IFG (p. Opercularis)	19.56
	AF3	F5	MFG	66.90
			IFG (p. Triangularis)	22.08
	F3	F1	MFG	68.06
			SFG	30.84
	F3	F5	MFG	60.23
			IFG (p. Triangularis)	38.27
	F3	FC3	MFG	81.08
			PG	9.36
	F7	FT7	IFG (p. Orbitalis)	27.4
			IFG (p. Triangularis)	25.14
	FC5	FT7	IFG (p. Opercularis)	36.59
			Rolandic Operculum	14.66
	FC5	FC3	Precentral Gyrus	46.40
			IFG (p. Opercularis)	18.71
Right	AF8	F6	IFG (p. Triangularis)	38.12
			MFG	37.98
	F8	F6	IFG (p. Triangularis)	73.88
			IFG (p. Orbitalis)	18.76
FC6	F6	IFG (p. Triangularis)	59.66	

Section 5: Role of prefrontal cortex in auditory short-term memory in children with dyslexia and developmental language disorder: a functional near-infrared spectroscopy study.

			IFG (p. Opercularis)	25.22
	AF4	F6	MFG	69.69
			SFG	14.98
	F4	F2	MFG	65.21
			SFG	31.75
	F4	F6	MFG	59.39
			IFG (p. Triangularis)	39.41
	F4	FC4	MFG	70.58
			IFG (p. Opercularis)	10.29
	F8	FT8	IFG (p. Orbitalis)	28.9
			Temporal Pole	28.9
	FC6	FT8	Rolandic operculum	25.6
			IFG (p. Opercularis)	23.8
	FC6	FC4	Precentral Gyrus	45.52
			IFG (p. Opercularis)	30.53

Table S1 : Each of the 20 recording channels is composed of a source and a detector (located in standard 10-20 positions). We report in the table the two cortical structures for which each channel has the highest specificity (in percent) according to the fOLD software (Zimeo Morais et al., 2018). SFG: Superior Frontal Gyrus, MFG: Middle Frontal Gyrus, IFG: Inferior Frontal Gyrus. The channels forming the IFG, the dlPFC (corresponding to MFG) ROIs are highlighted with light blue and light gold respectively. Channels unused in the present ROI analysis are not highlighted.

6 General Discussion

The discussion sections provided within each article presented in this PhD have already highlighted the significant implications of each study. The purpose of the current section is to establish connections between them. Firstly, we will discuss the general implications that this work has for the study of the development of auditory cognition. Secondly, we will discuss shared and distinct mechanisms at play in musical and verbal short-term memory (STM) by linking together the studies presented above and by presenting projects that I participated in during this PhD but that were not presented in the previous sections. Thirdly, we will discuss in more details the advantages and limitations of the functional near-infrared spectroscopy (fNIRS) method for the investigation of auditory cognition. Finally, we will present perspectives that this work brings for the diagnosis and remediation of learning disorders.

6.1 Development of auditory cognition

In the first two studies presented in the present PhD project, we explored the development of two fundamental auditory cognition processes: speech-in-noise perception (section 2) and auditory STM (section 3). Both processes seem to undergo a crucial maturation step around 7 years of age. At the anatomical level, three maturation periods of the human auditory system have been identified, with different developmental speed of maturation (for an exhaustive review, see Chapter 3 of Werner et al., 2012). The fastest maturation system concerns the cochlea and auditory nerve that reaches maturation around 3 months of age. The brainstem, including fibers in the medial geniculate body, reaches maturity around 1.5 years of age. The thalamus, auditory radiation and auditory cortex do not appear to reach maturity until 20 years of age. It thus appears that the structural underpinnings of the maturation step observed in our study probably lies in the

General discussion

cortical part of the auditory system. Notably, using MRI, Pujol et al. (2006) found that temporal cortex myelination is adult-like around 7-8 years of age. Using single photon emission computed tomography (SPECT) to explore resting-state regional cerebral blood flow (rCBF) in children, Devous et al. (2006) showed that secondary auditory area display a progressive reduction of rCBF with age (that the authors attribute to dendritic pruning processes) up until 19 years of age. However, primary auditory area (Heschl's gyrus) do not show this rCBF decrease with age from 7 years onward. These studies suggest that cortical primary auditory areas reach adult-like maturation around 7-8 years of age, while higher-level auditory area continue maturation up until adulthood. The maturation steps we observed in 7-year-olds in our studies can thus likely be attributed to anatomical maturation changes occurring in the primary auditory cortex.

This anatomical pivotal step of primary auditory area has found support in the observation of a maturation step around 7 years of age of low-level auditory processing. Indeed, Moore et al. (2011), in a series of three-interval, three-alternative (odd-one-out) forced choice tests, measured children's thresholds for frequency- and amplitude-modulation detection, frequency discrimination, and tone detection in quiet and various noise. Children of 6-7 years of age performed markedly lower than older children in all tasks, with a specific difficulty for the frequency discrimination task. These results suggest that auditory processing of low-level acoustic cues take a substantial rise after 7 years of age. Interestingly, children reached adult-like performance in all tests by 10-11 years of age, while in our study, children at 10 years of age did not reach adult performance, neither in speech-in-noise perception or auditory STM. This suggest that auditory processing of low-level cues cannot solely account for the developmental trajectory we observed.

General discussion

As mentioned above, while primary auditory areas reach maturation around 7 years of age, secondary and associative auditory area mature up until adulthood. Furthermore and as reviewed in section 1.2.1 and 1.2.2, speech processing in adverse conditions and auditory STM involve other regions than temporal regions, including the prefrontal cortex that, like secondary and associative auditory areas, do not reach maturity until 30 years of age (Kolk & Rakic, 2022). The prolonged developmental trajectory of higher-order cortical areas that are involved in the two auditory processes explored in section 2 and 3 might thus account for the absence of adult-like performance that we observed in the oldest children.

Another important factor that might have had an important influence on the developmental trajectory we observed for speech-in-noise perception and auditory STM, is the involvement of auditory selective attention in both processes. As mentioned in section 1.1.2.2, auditory selective attention plays a key role in the segregation of relevant auditory object against irrelevant ones and in WM/STM processes by (1) sustaining voluntary attention to the relevant auditory object and (2) efficiently ignoring auditory distractors. Importantly, Hoyer et al. (2021) showed that sustained attention abilities (as measured by increased response times variability and late response rate) do not reach adult-like performance before 14 years of age, while the ability to ignore distractors (as measured by missed targets when preceded by distractors) reaches maturity around 7-8 years of age. Together, these attentional factors could contribute to our observed 7 years old developmental step (distractibility) and absence of maturity by 10 years of age (sustained attention).

In summary, the common developmental trajectory we observed for speech-in-noise perception and both musical and verbal auditory STM in this PhD might rely on anatomical factors for which maturity seem to be reached in primary auditory areas at around 7 years of age, accounting for the observed low-level

auditory processing maturity and the ability to ignore distractors that seems to be reached at around this period of development. Later in development, the late-maturation of secondary auditory and prefrontal regions and the maturation of sustained attention abilities might account for the absence of adult-like performance at 10 years of age. Further studies should now screen speech-in-noise perception and auditory STM, in the same fashion as we did in the present PhD, but from 10 years of age until adulthood, a period during which we would expect a renewed increase of performance after puberty onset.

6.2 Musical and verbal STM

In the present PhD, three out of the four presented experimental sections concerned auditory STM for musical and verbal material and brought insights about the shared and distinct mechanisms between both materials. In the behavioral developmental study (section 3), we found several arguments for shared domain-general processes between musical and verbal STM: a similar developmental trajectory for both materials, similar recency effects, and significant correlations between STM performances of the two materials, probably due to shared attentional and serial-order processes in STM for both materials (see discussion section in section 3.2). In the fNIRS study in adults (section 4), we showed that, for both materials, lateral prefrontal regions (IFG and dlPFC) were engaged in auditory STM when compared to perceptual tasks and that the IFG responded parametrically to memory load increase for both materials, both results that underlie the fact that auditory STM processing occur in similar regions, as observed in previous fMRI studies.

Conversely, we uncovered evidence for distinct domain-specific processes. In section 3, we found that children performed poorer for the musical material than the verbal material for the same number of items and we found correlations between

General discussion

musical STM (but not verbal) and speech-in-noise perception performance. These results, in line with previously observed specific word/pitch-length, phonological/pitch proximity and irrelevant speech/pitch effects (see section 1.1.3.4) for each material, suggest that the sensory information of verbal and musical material are stored differently (i.e., different sensorimotor codes) for their maintenance and manipulation. Additionally, we uncovered specific hemodynamics for each material: while the IFG showed a parametric increase with memory load for verbal material, it showed a parametric decrease for musical material, suggesting that participants might change strategy (e.g., contour-based, chunking) to process musical sequences with higher memory load (see discussion section in section 4.2). These results are in line with Peretz & Zatorre (2005)'s hypothesis that while musical and verbal information share general WM/STM brain networks, these two types of information are not dynamically processed in the same way. Using fMRI, Albouy et al. (2019) brought evidence for this hypothesis by showing that amusics recruit similar high-level regions as controls during maintenance of verbal information in STM but show a decreased recruitment of right-lateralized structures for musical STM as compared to controls. As evidenced by the observed prefrontal hemodynamics recorded with fNIRS in our study, we brought further evidence that, while lateral prefrontal regions might be recruited by domain-general STM processes, their intrinsic dynamics might differ for handling different sensorimotor codes.

Additionally, in a project that was not reported in this PhD, in collaboration with Barbara Tillmann, Romain Quentin and Romaric Thiboud with data acquired and preprocessed by Yohana Lévêque and Lesly Fornoni, we explored auditory STM for musical and verbal material in 10 amusics and 10 controls that underwent a similar delayed matching-to-sample task (DMST) as the one used in section 3, 4, and 5 while being recorded with magnetoencephalography (MEG). For each

General discussion

participant, we used a linear classifier trained on the preprocessed raw signal of the 275 MEG sensors to decode the material (musical/verbal) that participants encoded during the first sequence of the DMST. By averaging the classifier's accuracy across the amusic and the control group, we observed that, in both groups, the classifier was able to decode above chance (up until 70% during the N100m; chance level: 50%) the material encoded by the participants. These results, in line with the hypothesis of different sensorimotor codes used to encode musical and verbal material, show that the neurophysiological signature during STM encoding of musical and verbal material is different enough for a classifier to be able to differentiate between materials. Additionally, the classifier decoded the material information above chance during the retention delay (after filtering data with a 1Hz low-pass filter), in keeping with the hypothesis of mechanisms of maintenance relying on domain-general processes common to the two materials, but acting on information based on different representations.

In another project started during a 4-month PhD mobility in the CERVO brain research center in Québec supervised by Philippe Albouy, we acquired stereo-encephalography (sEEG) data from drug-resistant epilepsy patients during their preoperative week while they underwent DMSTs for musical, verbal, and timbre sequences. By the insertion of thin electrodes directly into the brain, each recording at multiple sites, sEEG enables the capture of the electrophysiological activity of small neuron populations within the cortex with an excellent signal-to-noise ratio (Guenot et al., 2001). We used a custom-built convolutional neural network (CNN) created in the host laboratory to decode the material information (music/verbal/timbre) in each electrode. Preliminary results on 5 patients (84 electrodes) showed that the CNN was able to decode the material information above chance, during the encoding, the maintenance, and the retrieval of information in electrodes that recorded the auditory cortex and the IFG. These results provide

further evidence for differentiated neurophysiological signature of auditory STM for different material. Future steps include the recording of more data (currently 8 patients recorded), as, due to the heterogeneity of implantation schemas of each patient, more data is needed to ensure a full coverage of the cortical regions of interest. Moreover, further analysis will be performed with, notably, the decoding of item information per material (e.g., which word was encoded) and their mapping in the recorded cortical regions, decoding on time-frequency data, and correlations between the classifier's accuracy and participant's performance.

Overall, we brought evidence for the view that auditory STM for musical and verbal material, rather than being processed in the exact same way or in completely different systems, rely on both shared and distinct mechanisms. While common domain-general serial-order and attentional processes would be required to maintain information for both materials in lateral prefrontal regions preferentially (but not limited to them), their sensorimotor representations and the information on which maintenance strategies operate might result in different neurophysiological dynamics that fNIRS and machine-learning analysis methods seem to be able to uncover.

6.3 The fNIRS method: interests and limitations

In section 4 and 5 of this project, we used a promising new neuroimaging technique, fNIRS. As previously mentioned, fNIRS offer numerous advantages. First, its completely silent nature allowed us to record HbO and HbR concentration changes across the whole time-course of DMST trials. With fMRI, functional data during the whole-time course of a DMST trial is rarely obtained because of the noise generated by the machine when image slices are being obtained, constraining researchers to use sparse-sampling only when auditory sequences are being maintained, a disadvantage that is fully overcome with fNIRS. Moreover, fNIRS allows for a more precise functional characterization of the hemodynamics due to

its higher temporal sampling (7.81 data point per second in fNIRS against approximately one whole-brain image per second for a 3T fMRI). Second, the high signal-to-noise (SNR) ratio of the fNIRS signal and its low sensitivity to movement artifacts allows for a substantially lower number of trials than many electrophysiological or neuroimaging techniques. While several dozens of trials are usually necessary to observe event-related potentials/fields with EEG and MEG, we were able to observe our expected effects with only 12 trials per condition, thus drastically reducing testing time (even when considering the long trial and inter-trial durations). This is of utmost importance when testing children who can show rapid signs of irritability and distractibility when performing a fastidious task. We showed in section 5 that ~20 minutes testing time was perfectly feasible in our study, even by the youngest children (6 years old). Third, fNIRS hardware is less constraining to install on participant's scalp than EEG ones. In the latter, the use of conductive gel and the number of electrodes to install can render the installation particularly time-consuming and uncomfortable for participants, especially for children. In our studies, the time we spent installing the fNIRS cap never exceeded 25 minutes in adults and 17 minutes in children. It should however be noted that the fNIRS cap and the optode contact on the scalp can become uncomfortable after a certain amount of time, a constraint that is counterbalanced by fNIRS' previously mentioned high SNR that allows shorter testing times. Fourth, despite of a lack of definite consensus, preprocessing of fNIRS signal is relatively straightforward (see section 1.2.3.4), due to its low sensitivity to movements and its high SNR. In summary, fNIRS brings distinctive advantages to the field of neuroscience and we showed, in section 4 and 5, that it was particularly advantageous in the context of our research. We will now outline several challenges we encountered while recording, preprocessing, and analyzing fNIRS data. Our intention is to offer

insights that could prove valuable to both the fNIRS community and researchers who are new at employing this technique.

One of the limitations of this technique concerns the constraints on the inclusion of certain participants, due to the physical properties of the near-infrared light (NIR). Optical neuroimaging techniques heavily depend on the scattering and absorption of NIR light in human tissue, which is influenced by factors like the concentration of chromophores such as melanin (Bashkatov et al., 2011). Consequently, noise levels are consistently higher in individuals with darker skin and hair pigmentation due to increased light absorption. Additionally, frizzy or afro-textured hair, as well as protective hairstyles, can jeopardize the access to secure direct and consistent optode-to-scalp contact needed for measurements, similar to EEG recordings. Although these issues are usually recognized as methodological limitations rather than essential equity concerns, they result in skin-tone and hair-style biases that exclude a significant portion of the population, namely people with black and brown skin. The exclusion of racially and ethnically marginalized individuals due to perceived “unusable” data aligns with broader structural injustices faced by black and brown communities worldwide (Galán et al., 2021). Even with no explicit intention from the researcher to generate biased results, research tools and protocols often perpetuate the generation of scientific results that disproportionately center on a specific demographic, reinforcing white-normative standards (Muthukrishna et al., 2020). While reviewing structural injustices towards minority groups in science is beyond the scope of this discussion, we thought that this particular issue encountered with fNIRS was worth mentioning, as it tends to perpetuate such structural issues. Care should be taken in the way we gather data and report scientific results as they tend to be perceived, by their nature, as unbiased information, when in fact, they can substantially reinforce social injustice (Webb et al., 2022).

General linear models (GLM) are proving to be a powerful analysis tool for fNIRS data, as it does for fMRI, both in terms of its advantages regarding the less constraining event-related experimental design that can be used and its ability to deconvolute the hemodynamic response for short-separated events (see section 1.2.3.3 and 1.2.3.5). With this approach, the entirety of the recording session must be fed to the statistical model, making it mandatory to acquire data in one recording file. However, during a recording session, breaks can be necessary to give instructions to participants or to give participants the opportunity to rest for a few minutes. These moments, when the participant is not engaged in the task and generally produces a lot of parasitic movements, should not be fed to the GLM. To overcome this obstacle, two possibilities arise: either allow the introduction of breaks during a single recording session in the fNIRS acquisition software, or implement functions in analysis software that allow the trimming of these breaks from the acquired raw data during the first preprocessing steps. In our case, both solutions were unavailable and we were constrained to build our own trimming functions. Even if such considerations can seem trivial and easily overcome, solving these problems, in the absence of solutions already implemented in acquisition and analysis software, can quickly become time-consuming.

Another challenge we faced was the lack of standardization of preprocessing and analysis of fNIRS data, and in particular the management of systemic physiological responses. As reviewed in section 1.2.3.4, spontaneous vasomotor components for which oscillations fall into the frequency range of interest (e.g., stimulus frequency and hemodynamic response rate, ~ 0.1 Hz) cannot be excluded with a standard band-pass filtering method. Cleaning data from these components have been shown to be challenging with methods such as principal/independent component analysis (P/ICA) and the increasingly preferred approach is to incorporate short-separation (SS) channels data directly into a GLM as regressors.

It has been showed that the best method was to incorporate all SS data (HbO and HbR) in the GLM to remove physiological data from each channel of interest (Santosa et al., 2020), a method we used in section 4 and 5. Indeed, in our dataset, we compared informally the results of our analysis with and without SS regression, and results differed drastically between the two analysis, sometimes even showing opposite effects from the one we expected when SS data were not regressed. When we looked closer at the systemic signal recorded by SS channels, the signal seemed to respond specifically to the stimulation, showing that vasomotor systemic signal is not independent from the stimulation, and should thus always be considered in the analysis. Fortunately, SS channels have become available in most fNIRS device nowadays, but this raises questions about fNIRS studies that did not perform any cleaning (with SS channels or other methods) of these components, as these systemic components appear to have a significant impact on the obtained results. Additionally, we noticed that there is a lack of available functions in the existing analysis software to regress SS data from the preprocessed fNIRS signal. Namely, when one wants to average the HbO/HbR change around an event of interest (averaging approach, see section 1.2.3.5), there is a lack of option and consensus on how to regress SS data from fNIRS signal. In our case, we wanted to average our fNIRS data as a first intention in order to familiarize ourselves with the signal and explore the HbO/HbR concentration changes before it was processed by a statistical model. While one analysis toolbox provided a way to regress some of the SS data from HbO/HbR concentration signal, no option was available to include the entirety of SS data in the regression. We thus regressed SS data from the fNIRS signal using a GLM from a custom-written function. As for the trimming of data mentioned above, implementing such methods in the analysis software would avoid significantly time-consuming steps and allow for an easier exploration of fNIRS data.

Finally, despite substantial efforts to establish a universal data format (e.g., shared near infrared spectroscopy format; SNIRF), existing software currently lack compatibility between their own specific data formats, making it challenging to leverage the unique advantages of each software. Furthermore, the novelty of the technique results in inadequate documentation for some of these software options.

In conclusion, fNIRS presents a valuable tool in the field of neuroscience due to its non-invasiveness, portability, and suitability to explore the auditory modality. The methodological flexibility and suitability for various populations, including children and clinical groups, underscore its wide applicability. However, it is essential to acknowledge its limitations, including challenges in the recruitment of heterogeneous populations, the absence of versatile analysis tools and the ongoing efforts required to standardize preprocessing and analysis procedures. Despite these limitations, fNIRS contributes significantly to our understanding of brain function, paving the way for further advancements, in particular regarding our understanding of language-related learning disorders.

6.4 Perspectives for learning disorders

Most learning disorders are currently diagnosed with reading tests, meaning that children have to have acquired literacy to be included in a personalized learning program. However, these children often go through repeated difficulties in the acquisition of language and reading all along the years prior to their diagnosis. As reviewed in section 1.1, speech-in-noise perception, auditory attention, and auditory STM are consistently impaired in dyslexia and developmental learning disorders (DLD). In the present PhD, we showed that among the five children who presented a diagnosis of dyslexia out of the 100 children tested in section 2 and 3, all of them showed lower performance than the mean of other children for musical STM and the easy noise/phonological conditions of the speech-in-noise test, and

four of them presented lower performance in verbal STM and the harder conditions of the speech-in-noise test. We also showed that both tasks did not require reading acquisition and that children as young as 5 years of age (before reading acquisition), were able to perform the tasks, making them particularly suited for the screening of pre-reading children. This stresses the necessity to identify markers that could inform us about the potential future difficulties that a pre-reading child could encounter later in development. Thus, the screening of auditory cognition in children stands out as a relevant candidate for the identification of such markers.

Currently in France, 3-4 years old and 6 years old children undergo a mandatory medical check-up (Article L541-1 of the “Code de la Santé Publique”). Although the official legal texts stipulate the obligation for national education health services to screen children for potential peripheral hearing disorders and language-related learning disorders, no standardized procedure is provided and the administration of screening tests is left to the discretion of the mandated health professional. Consequently, none of these medical check-ups include central auditory processing screening tests, and screening tests assessing learning disorders usually only include reading tests. Therefore, we argue that incorporating screening tests that evaluate central auditory processes would represent a significant advancement in the early identification and intervention for children exhibiting deficits in the aforementioned auditory cognitive processes. This proactive approach could potentially prevent the numerous social and emotional challenges that children encounter prior to receiving a correct diagnosis.

The identification of objective markers of language-related learning disorders seem crucial to complement behavioral screening in order to uncover the neural activity associated with central auditory processes. For example, as children with learning disorders grow older, they tend to compensate for tasks that the education system considers mandatory to master during language and reading development

(Van Viersen et al., 2019). In consequence, while some of them could display behavioral outputs similar to typically developing (TD) children, the effort necessary for children with learning disorders to produce such an output could be drastically higher than for control children. In the case of auditory STM, we showed in section 4 that fNIRS was a valuable tool to uncover neural activity associated with increasing difficulty (i.e., memory load), and preliminary results in section 5 suggest that we will be able to uncover these mechanisms in school-aged children. Crucially, in addition with the group analysis planned for this study (see Discussion in section 5.2), we will be able to compare fNIRS results from children with learning disorders who have compensated their impairments (as measured by the standardized reading tests that they underwent during the screening session and their results from the auditory STM tasks) and compare them to those of TD children. In fact, increased activation in lateral prefrontal regions have been suggested to be involved in compensatory mechanisms of reading difficulties (D’Mello & Gabrieli, 2018; Hoeft et al., 2007; Price, 2012). In the hypothesis that children with learning disorders recruit the same lateral prefrontal regions as TD children to perform the auditory STM task, we should observe high level of activity already for small memory load, leading to an absence of parametric activation with increased memory load, in contrast with TD children. Such results would confirm that despite similar behavioral outcomes, children with learning disorder engage in higher cognitive effort than their TD peers to perform similar tasks.

In summary, exploring central auditory processes in children with language-related learning disorders is of paramount importance. Behavioral central auditory screening methods should be systematically included in children’s medical check-ups as they don’t rely on reading acquisition and could thus function as efficient proxy of future diagnosis of learning disorders. In addition, the use of portable, non-invasive and child-friendly neuroimaging technique such as fNIRS allows to

General discussion

delve deeper into the physiological underpinnings of auditory cognition and potentially identify early markers of language-related learning disorders. This comprehensive approach holds the potential to enhance early identification, intervention, and ultimately improve the quality of life for these children by addressing their unique cognitive needs more effectively.

Conclusions

The present PhD work has brought new insights on the development of auditory cognition during school years by integrating behavioral measures of two fundamental auditory processes, speech-in-noise perception and auditory STM, and the identification of objective neurophysiological markers of auditory STM using fNIRS, a promising neuroimaging technique. Our results have brought insights into the developmental trajectory of these two auditory processes and to the proposition that they undergo a long process of maturation that extends until adolescence and that a decisive developmental step occurs around 7 years of age. Moreover, we brought insights about the importance of investigating auditory STM for musical and verbal material, as our results suggest that distinct processes between the two materials might underlie their representations in STM and that shared processes might contribute to the central auditory processing deficits often observed in language-related learning disorders. We also evidenced the well-suitedness of fNIRS for exploring auditory STM and for uncovering the complex neural dynamics subtending the processing of musical and verbal material in STM. Finally, we showed promising preliminary results and perspectives that fNIRS brings to the exploration of auditory STM deficits in children with language-related learning disorders. In conclusion, we hope that this work will help to highlight the importance of central auditory processes in children's social and educational lives, and that it will motivate the relatively young fNIRS community to continue exploring auditory cognition as a whole, and in particular in children with language-related learning disorders.

References

- Aben, B., Stapert, S., & Blokland, A. (2012). About the Distinction between Working Memory and Short-Term Memory. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00301>
- Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, 50(1), 77-84. <https://doi.org/10.1016/j.neuropsychologia.2011.10.024>
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension : Adapting to time-compressed speech. *NeuroImage*, 49(1), 1124-1132. <https://doi.org/10.1016/j.neuroimage.2009.07.032>
- Ahmed, S. F., Ellis, A., Ward, K. P., Chaku, N., & Davis-Kean, P. E. (2022). Working memory development from early childhood to adolescence using two nationally representative samples. *Developmental Psychology*, 58(10), 1962-1973. <https://doi.org/10.1037/dev0001396>
- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). “What” and “where” in the human auditory system. *Proceedings of the National Academy of Sciences*, 98(21), 12301-12306. <https://doi.org/10.1073/pnas.211209098>
- Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, 367(6481), 1043-1047. <https://doi.org/10.1126/science.aaz3468>
- Albouy, P., Caclin, A., Norman-Haignere, S. V., Lévêque, Y., Peretz, I., Tillmann, B., & Zatorre, R. J. (2019). Decoding Task-Related Functional Brain Imaging Data to Identify Developmental Disorders : The Case of Congenital Amusia. *Frontiers in Neuroscience*, 13, 1165. <https://doi.org/10.3389/fnins.2019.01165>
- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia : The deficit starts in the auditory cortex. *Brain*, 136(5), 1639-1661. <https://doi.org/10.1093/brain/awt082>
- Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory : fMRI evidence in congenital amusia. *Human Brain Mapping*, 40(3), 855-867. <https://doi.org/10.1002/hbm.24416>
- Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, 1537, 224-232. <https://doi.org/10.1016/j.brainres.2013.09.003>

References

- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, 55(1), 64-88.
<https://doi.org/10.1016/j.jml.2006.02.002>
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children : Are They Separable? *Child Development*, 77(6), 1698-1716.
<https://doi.org/10.1111/j.1467-8624.2006.00968.x>
- Anourova, I., Nikouline, V. V., Ilmoniemi, R. J., Hotta, J., Aronen, H. J., & Carlson, S. (2001). Evidence for Dissociation of Spatial and Nonspatial Auditory Information Processing. *NeuroImage*, 14(6), 1268-1277.
<https://doi.org/10.1006/nimg.2001.0903>
- Aslin, R. N. (2012). Questioning the questions that have been asked about the infant brain using near-infrared spectroscopy. *Cognitive Neuropsychology*, 29(1-2), 7-33. <https://doi.org/10.1080/02643294.2012.654773>
- Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants : Promise, prospects, and challenges. *Journal of Biomedical Optics*, 10(1), 011009.
<https://doi.org/10.1117/1.1854672>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory : A Proposed System and its Control Processes. In *Psychology of Learning and Motivation* (Vol. 2, p. 89-195). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppel, R. A., & Katz, S. (1996). Dissociation of Storage and Rehearsal in Verbal Working Memory : Evidence From Positron Emission Tomography. *Psychological Science*, 7(1), 25-31. <https://doi.org/10.1111/j.1467-9280.1996.tb00662.x>
- Baddeley, A. D. (1966). The Influence of Acoustic and Semantic Similarity on Long-term Memory for Word Sequences. *Quarterly Journal of Experimental Psychology*, 18(4), 302-309.
<https://doi.org/10.1080/14640746608400047>
- Baddeley, A. D. (1992). Working Memory. *Science*, 255(5044), 556-559.
<https://doi.org/10.1126/science.1736359>
- Baddeley, A. D. (1996). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, 49(1), 5-28.
<https://doi.org/10.1080/713755608>
- Baddeley, A. D. (2000). The episodic buffer : A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.
[https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D. (2010). Working memory. *Current Biology*, 20(4), R136-R140.
<https://doi.org/10.1016/j.cub.2009.12.014>

References

- Baddeley, A. D. (2012). Working Memory : Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1-29.
<https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation*, 8, 47-90.
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586-595. [https://doi.org/10.1016/0749-596X\(88\)90028-9](https://doi.org/10.1016/0749-596X(88)90028-9)
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575-589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)
- Ballot, C., Mathey, S., & Robert, C. (2021). Word imageability and orthographic neighbourhood effects on memory : A study in free recall and recognition. *Memory*, 29(6), 829-834. <https://doi.org/10.1080/09658211.2021.1921216>
- Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory Distraction and Short-Term Memory : Phenomena and Practical Implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(1), 12-29.
<https://doi.org/10.1518/001872001775992462>
- Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Éds.), *The Cognitive Neuroscience of Working Memory* (p. 59-80). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198570394.003.0004>
- Bashkatov, A. N., Genina, E. A., & Tuchin, V. V. (2011). Optical Properties Of Skin, Subcutaneous, And Muscle Tissues : A Review. *Journal of Innovative Optical Health Sciences*, 04(01), 9-38.
<https://doi.org/10.1142/S1793545811001319>
- Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of Auditory-evoked Potentials in Musicians Reflects an Influence of Expertise but not Selective Attention. *Journal of Cognitive Neuroscience*, 20(12), 2238-2249.
<https://doi.org/10.1162/jocn.2008.20157>
- Beauchamp, M. S., Beurlot, M. R., Fava, E., Nath, A. R., Parikh, N. A., Saad, Z. S., Bortfeld, H., & Oghalai, J. S. (2011). The Developmental Trajectory of Brain-Scalp Distance from Birth through Childhood : Implications for Functional Neuroimaging. *PLoS ONE*, 6(9), e24981.
<https://doi.org/10.1371/journal.pone.0024981>
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161-1165.
<https://doi.org/10.1038/nature03867>

References

- Berz, W. L. (1995). Working Memory in Music : A Theoretical Model. *Music Perception: An Interdisciplinary Journal*, 12(3), 353-364.
<https://doi.org/10.2307/40286188>
- Binder, J. R. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, 10(5), 512-528.
<https://doi.org/10.1093/cercor/10.5.512>
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3), 295-301.
<https://doi.org/10.1038/nn1198>
- Blain, S., Talamini, F., Fornoni, L., Bidet-Caulet, A., & Caclin, A. (2022). Shared cognitive resources between memory and attention during sound-sequence encoding. *Attention, Perception, & Psychophysics*, 84(3), 739-759.
<https://doi.org/10.3758/s13414-021-02390-2>
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and Verbal Memory Span : A Meta-Analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(5), P223-P233.
<https://doi.org/10.1093/geronb/60.5.P223>
- Botvinick, M. M. (2005). Effects of domain-specific knowledge on memory for serial order. *Cognition*, 97(2), 135-151.
<https://doi.org/10.1016/j.cognition.2004.09.007>
- Botvinick, M. M., & Bylsma, L. M. (2005). Regularization in Short-Term Memory for Serial Order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 351-358. <https://doi.org/10.1037/0278-7393.31.2.351>
- Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature*, 271(5647), 746-747.
<https://doi.org/10.1038/271746a0>
- Braukmann, R., Lloyd-Fox, S., Blasi, A., Johnson, M. H., Bekkering, H., Buitelaar, J. K., & Hunnius, S. (2018). Diminished socially selective neural processing in 5-month-old infants at high familial risk of autism. *European Journal of Neuroscience*, 47(6), 720-728. <https://doi.org/10.1111/ejn.13751>
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *NeuroImage*, 5(1), 49-62.
<https://doi.org/10.1006/nimg.1996.0247>
- Bregman, A. S. (1994). *Auditory scene analysis : The perceptual organization of sound*. MIT press.
- Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., & Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy : A comparison of motion correction

References

- techniques applied to real cognitive data. *NeuroImage*, 85, 181-191.
<https://doi.org/10.1016/j.neuroimage.2013.04.082>
- Brigadoi, S., & Cooper, R. J. (2015). How short is short? Optimum source–detector distance for short-separation channels in functional near-infrared spectroscopy. *Neurophotonics*, 2(2), 025005.
<https://doi.org/10.1117/1.NPh.2.2.025005>
- Bright, R. (1831). *Diseases of the brain and nervous system (Reports of medical cases selected with a view of illustrating the symptoms and care of diseases by a reference to morbid anatomy, p. 431-435)*.
- Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.
<https://doi.org/10.1037/10037-000>
- Brown, D. K., Cameron, S., Martin, J. S., Watson, C., & Dillon, H. (2010). The North American Listening in Spatialized Noise—Sentences Test (NA LiSN-S) : Normative Data and Test-Retest Reliability Studies for Adolescents and Young Adults. *Journal of the American Academy of Audiology*, 21(10), 629-641. <https://doi.org/10.3766/jaaa.21.10.3>
- Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults : Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America*, 141(4), 2650-2660.
<https://doi.org/10.1121/1.4979936>
- Butler, L. K., Kiran, S., & Tager-Flusberg, H. (2020). Functional Near-Infrared Spectroscopy in the Study of Speech and Language Impairment Across the Life Span : A Systematic Review. *American Journal of Speech-Language Pathology*, 29(3), 1674-1701. https://doi.org/10.1044/2020_AJSLP-19-00050
- Cabbage, K. L., & Hitchcock, E. R. (2022). Clinical Considerations for Speech Perception in School-Age Children With Speech Sound Disorders : A Review of the Current Literature. *Language, Speech, and Hearing Services in Schools*, 53(3), 768-785. https://doi.org/10.1044/2022_LSHSS-21-00120
- Caclin, A. (2021). *Déficits et remédiation de la cognition auditive chez l'Humain : Aspects comportementaux et neurophysiologiques [Habilitation à Diriger des Recherches]*. https://pam-lyon.cnrs.fr/wp-content/uploads/2021/06/HDR_ACaclin_Final.pdf
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory : Insights from neurodevelopmental and neurological disorders: Musical and verbal short-term memory. *Annals of the New York Academy of Sciences*, 1423(1), 155-165. <https://doi.org/10.1111/nyas.13733>
- Calandruccio, L., Leibold, L. J., & Buss, E. (2016). Linguistic Masking Release in School-Age Children and Adults. *American Journal of Audiology*, 25(1), 34-40. https://doi.org/10.1044/2015_AJA-15-0053

References

- Calcus, A., Deltenre, P., Colin, C., & Kolinsky, R. (2018). Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Developmental Science*, 21(3), e12558.
<https://doi.org/10.1111/desc.12558>
- Camina, E., & Güell, F. (2017). The Neuroanatomical, Neurophysiological and Psychological Basis of Memory : Current Models and Their Origins. *Frontiers in Pharmacology*, 8, 438.
<https://doi.org/10.3389/fphar.2017.00438>
- Catani, M., & de Schotten, T. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8), 1105-1132.
<https://doi.org/10.1016/j.cortex.2008.05.004>
- Chakravarti, S., Srivastava, S., & Mittnacht, A. J. C. (2008). Near Infrared Spectroscopy (NIRS) in Children. *Seminars in Cardiothoracic and Vascular Anesthesia*, 12(1), 70-79. <https://doi.org/10.1177/1089253208316444>
- Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature*, 396(6707), 128-128. <https://doi.org/10.1038/24075>
- Chance, B., Zhuang, Z., UnAh, C., Alter, C., & Lipton, L. (1993). Cognition-activated low-frequency modulation of light absorption in human brain. *Proceedings of the National Academy of Sciences*, 90(8), 3770-3774.
<https://doi.org/10.1073/pnas.90.8.3770>
- Chang, S.-E. (2014). Research Updates in Neuroimaging Studies of Children Who Stutter. *Seminars in Speech and Language*, 35(02), 067-079.
<https://doi.org/10.1055/s-0034-1382151>
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25(5), 975-979. <https://doi.org/10.1121/1.1907229>
- Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and Spatial Short-Term Memory : Common Sources of Developmental Change? *Journal of Experimental Child Psychology*, 73(1), 7-44.
<https://doi.org/10.1006/jecp.1999.2493>
- Clark, K. M., Hardman, K. O., Schachtman, T. R., Saults, J. S., Glass, B. A., & Cowan, N. (2018). Tone series and the nature of working memory capacity development. *Developmental Psychology*, 54(4), 663-676.
<https://doi.org/10.1037/dev0000466>
- Cohen, Y. E., Popper, A. N., & Fay, R. R. (Éds.). (2013). *Neural Correlates of Auditory Cognition* (Vol. 45). Springer New York.
<https://doi.org/10.1007/978-1-4614-2350-8>
- Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., & Boas, D. A. (2012). A Systematic Comparison of Motion Artifact Correction Techniques for Functional Near-Infrared Spectroscopy. *Frontiers in Neuroscience*, 6. <https://doi.org/10.3389/fnins.2012.00147>

References

- Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of Open-Set Word Recognition in Children : Speech-Shaped Noise and Two-Talker Speech Maskers. *Ear & Hearing, 37*(1), 55-63.
<https://doi.org/10.1097/AUD.0000000000000201>
- Costers, L., Van Schependom, J., Laton, J., Baijot, J., Sjøgård, M., Wens, V., De Tiège, X., Goldman, S., D'Haeseleer, M., D'hooghe, M. B., Woolrich, M., & Nagels, G. (2020). Spatiotemporal and spectral dynamics of multi-item working memory as revealed by the n -back task using MEG. *Human Brain Mapping, 41*(9), 2431-2446. <https://doi.org/10.1002/hbm.24955>
- Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia, 155*, 107811.
<https://doi.org/10.1016/j.neuropsychologia.2021.107811>
- Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology, 36*(1-2), 1-17.
<https://doi.org/10.1080/02643294.2019.1578205>
- Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology, 1-47*.
<https://doi.org/10.1080/09297049.2022.2162031>
- Cowan, N. (1984). On Short and Long Auditory Stores. *Psychological Bulletin, 96*(2), 341-370. <https://doi.org/10.1037/0033-2909.96.2.341>
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163-191.
<https://doi.org/10.1037/0033-2909.104.2.163>
- Cowan, N. (1998). *Attention and Memory : An Integrated Framework*.
- Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research* (Vol. 169, p. 323-338). Elsevier. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 599-607.
[https://doi.org/10.1016/S0022-5371\(73\)80039-8](https://doi.org/10.1016/S0022-5371(73)80039-8)
- Cristia, A., Dupoux, E., Hakuno, Y., Lloyd-Fox, S., Schuetze, M., Kivits, J., Bergvelt, T., Van Gelder, M., Filippin, L., Charron, S., & Minagawa-Kawai, Y. (2013). An Online Database of Infant Functional Near InfraRed Spectroscopy Studies : A Community-Augmented Systematic Review. *PLoS ONE, 8*(3), e58906. <https://doi.org/10.1371/journal.pone.0058906>

References

- Croonen, W. L. M. (1994). Effects of length, tonal structure, and contour in the recognition of tone series. *Perception & Psychophysics*, 55(6), 623-632. <https://doi.org/10.3758/BF03211677>
- Csibra, G., Tucker, L. A., & Johnson, M. H. (2001). Differential Frontal Cortex Activation Before Anticipatory and Reactive Saccades in Infants. *Infancy*, 2(2), 159-174. https://doi.org/10.1207/S15327078IN0202_3
- Cutini, S., Szűcs, D., Mead, N., Huss, M., & Goswami, U. (2016). Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia. *NeuroImage*, 143, 40-49. <https://doi.org/10.1016/j.neuroimage.2016.08.012>
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, 23(8), 3423-3431. <https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003>
- Delpy, D. T., Cope, M., Zee, P. V. D., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, 33(12), 1433-1442. <https://doi.org/10.1088/0031-9155/33/12/008>
- D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 761-772. <https://doi.org/10.1098/rstb.2007.2086>
- Deutsch, D. (1970). Tones and Numbers: Specificity of Interference in Immediate Memory. *Science*, 168(3939), 1604-1605. <https://doi.org/10.1126/science.168.3939.1604>
- Deutsch, D. (1974). Generality of Interference by Tonal Stimuli in Recognition Memory for Pitch. *Quarterly Journal of Experimental Psychology*, 26(2), 229-234. <https://doi.org/10.1080/14640747408400408>
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 80-90. <https://doi.org/10.1037/h0039515>
- Devous, M. D., Altuna, D., Furl, N., Cooper, W., Gabbert, G., Ngai, W. T., Chiu, S., Scott, J. M., Harris, T. S., Payne, J. K., & Tobey, E. A. (2006). Maturation of Speech and Language Functional Neuroanatomy in Pediatric Normal Controls. *Journal of Speech, Language, and Hearing Research*, 49(4), 856-866. [https://doi.org/10.1044/1092-4388\(2006/061\)](https://doi.org/10.1044/1092-4388(2006/061))
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, 109(8). <https://doi.org/10.1073/pnas.1113427109>
- D'Mello, A. M., & Gabrieli, J. D. E. (2018). Cognitive Neuroscience of Dyslexia. *Language, Speech, and Hearing Services in Schools*, 49(4), 798-809. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0020
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in*

References

- Cognitive Sciences, 12(3), 99-105.
<https://doi.org/10.1016/j.tics.2008.01.001>
- Dowell, F. E., Throne, J. E., Wang, D., & Baker, J. E. (1999). Identifying Stored-Grain Insects Using Near-Infrared Spectroscopy. *Journal of Economic Entomology*, 92(1), 165-169. <https://doi.org/10.1093/jee/92.1.165>
- Doyle, A.-B. (1973). Listening to distraction : A developmental study of selective attention. *Journal of Experimental Child Psychology*, 15(1), 100-115.
[https://doi.org/10.1016/0022-0965\(73\)90134-3](https://doi.org/10.1016/0022-0965(73)90134-3)
- Drachman, D. A. (1966). Memory and the Hippocampal Complex : II. Is Memory a Multiple Process? *Archives of Neurology*, 15(1), 52.
<https://doi.org/10.1001/archneur.1966.00470130056005>
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92(1), 53-78.
<https://doi.org/10.1348/000712601162103>
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87(3), 272-300.
<https://doi.org/10.1037/0033-295X.87.3.272>
- Ebbinghaus, H. (2013). *Memory : A Contribution to Experimental Psychology*. *Annals of Neurosciences*, 20(4).
<https://doi.org/10.5214/ans.0972.7531.200408>
- Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear & Hearing*, 37(1), 101S-110S.
<https://doi.org/10.1097/AUD.0000000000000300>
- Eckert, M. A., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2008). Age-related Effects on Word Recognition : Reliance on Cognitive Control Systems with Structural Declines in Speech-responsive Cortex. *Journal of the Association for Research in Otolaryngology*, 9(2), 252-259. <https://doi.org/10.1007/s10162-008-0113-3>
- Elliott, B. L., & Brewer, G. A. (2019). Divided Attention Selectively Impairs Value-Directed Encoding. *Collabra: Psychology*, 5(1), 4.
<https://doi.org/10.1525/collabra.156>
- Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *The Journal of the Acoustical Society of America*, 66(3), 651-653. <https://doi.org/10.1121/1.383691>
- Emberson, L. L., Zinszer, B. D., Raizada, R. D. S., & Aslin, R. N. (2017). Decoding the infant mind : Multivariate pattern analysis (MVPA) using fNIRS. *PLOS ONE*, 12(4), e0172500.
<https://doi.org/10.1371/journal.pone.0172500>

References

- Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1), 19-23.
<https://doi.org/10.1111/1467-8721.00160>
- Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence : A Latent-Variable Approach.
- Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2), 921-935.
<https://doi.org/10.1016/j.neuroimage.2012.03.049>
- Ferreri, L., Bigand, E., Perrey, S., & Bugajska, A. (2014). The promise of Near-Infrared Spectroscopy (NIRS) for psychological research : A brief review. *L'Année Psychologique*, 114(03), 537-569.
<https://doi.org/10.4074/S0003503314003054>
- Filippetti, M. L., Andreu-Perez, J., de Klerk, C., Richmond, C., & Rigato, S. (2023). Are advanced methods necessary to improve infant fNIRS data analysis? An assessment of baseline-corrected averaging, general linear model (GLM) and multivariate pattern analysis (MVPA) based approaches. *NeuroImage*, 265, 119756.
<https://doi.org/10.1016/j.neuroimage.2022.119756>
- Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal Derivative Distribution Repair (TDDR) : A motion correction method for fNIRS. *NeuroImage*, 184, 171-179.
<https://doi.org/10.1016/j.neuroimage.2018.09.025>
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract : A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511-1522.
<https://doi.org/10.1121/1.427148>
- Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental Effects in Children's Ability to Benefit From F0 Differences Between Target and Masker Speech. *Ear & Hearing*, 40(4), 927-937.
<https://doi.org/10.1097/AUD.0000000000000673>
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. *Child Development*, 37(2), 283. <https://doi.org/10.2307/1126804>
- Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., & Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nature human behaviour*, 3(4), 393-405.
<https://doi.org/10.1038/s41562-019-0548-z>
- Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-

References

- reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, 25(4), 383-390.
<https://doi.org/10.1525/mp.2008.25.4.383>
- Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*, 75, 27-35. <https://doi.org/10.1016/j.neuroimage.2013.02.044>
- Foster, N. E. V., & Zatorre, R. J. (2010). A Role for the Intraparietal Sulcus in Transforming Musical Pitch Information. *Cerebral Cortex*, 20(6), 1350-1359. <https://doi.org/10.1093/cercor/bhp199>
- Fox, S. E., Wagner, J. B., Shrock, C. L., Tager-Flusberg, H., & Nelson, C. A. (2013). Neural Processing of Facial Identity and Emotion in Infants at High-Risk for Autism Spectrum Disorders. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00089>
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science*, 291(5502), 312-316. <https://doi.org/10.1126/science.291.5502.312>
- Frey, S., Campbell, J. S. W., Pike, G. B., & Petrides, M. (2008). Dissociating the Human Language Pathways with High Angular Resolution Diffusion Fiber Tractography. *The Journal of Neuroscience*, 28(45), 11435-11444.
<https://doi.org/10.1523/JNEUROSCI.2388-08.2008>
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246-2256. <https://doi.org/10.1121/1.1689343>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189-210.
<https://doi.org/10.1002/hbm.460020402>
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*, 17(4), 437-455. <https://doi.org/10.1016/j.conb.2007.07.011>
- Fu, G., Wan, N. J. A., Baker, J. M., Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2016). A Proof of Concept Study of Function-Based Statistical Analysis of fNIRS Data: Syntax Comprehension in Children with Specific Language Impairment Compared to Typically-Developing Controls. *Frontiers in Behavioral Neuroscience*, 10.
<https://doi.org/10.3389/fnbeh.2016.00108>
- Fu, X., & Richards, J. E. (2021). Investigating developmental changes in scalp-to-cortex correspondence using diffuse optical tomography sensitivity in infancy. *Neurophotonics*, 8(03). <https://doi.org/10.1117/1.NPh.8.3.035003>

References

- Funabiki, Y., Murai, T., & Toichi, M. (2012). Cortical activation during attention to sound in autism spectrum disorders. *Research in Developmental Disabilities, 33*(2), 518-524. <https://doi.org/10.1016/j.ridd.2011.10.016>
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—An fMRI study with sparse temporal sampling. *NeuroImage, 19*(4), 1417-1426. [https://doi.org/10.1016/S1053-8119\(03\)00224-6](https://doi.org/10.1016/S1053-8119(03)00224-6)
- Galaburda, A. M. (1978). Human Brain : Cytoarchitectonic Left-Right Asymmetries in the Temporal Speech Region. *Archives of Neurology, 35*(12), 812. <https://doi.org/10.1001/archneur.1978.00500360036007>
- Galán, C. A., Bekele, B., Boness, C., Bowdring, M., Call, C., Hails, K., McPhee, J., Mendes, S. H., Moses, J., Northrup, J., Rupert, P., Savell, S., Sequeira, S., Tervo-Clemmens, B., Tung, I., Vanwoerden, S., Womack, S., & Yilmaz, B. (2021). Editorial: A Call to Action for an Antiracist Clinical Science. *Journal of Clinical Child & Adolescent Psychology, 50*(1), 12-57. <https://doi.org/10.1080/15374416.2020.1860066>
- Gathercole, S. E. (1998). The Development of Memory. *Journal of Child Psychology and Psychiatry, 39*(1), 3-27. <https://doi.org/10.1017/S0021963097001753>
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419. [https://doi.org/10.1016/S1364-6613\(99\)01388-1](https://doi.org/10.1016/S1364-6613(99)01388-1)
- Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*(4), 770-778. <https://doi.org/10.1037/0012-1649.29.4.770>
- Gathercole, S. E., Adams, A.-M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition, 22*(2), 201-207. <https://doi.org/10.3758/BF03208891>
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 84-95. <https://doi.org/10.1037/0278-7393.25.1.84>
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology, 40*(2), 177-190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable Lexical and Phonological Influences on Serial Recognition and Serial Recall. *The Quarterly Journal of Experimental Psychology Section A, 54*(1), 1-30. <https://doi.org/10.1080/02724980042000002>
- Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5- to 10-year-old children :

References

- Focus on musical and verbal short-term memory. *Developmental Science*, 25(3). <https://doi.org/10.1111/desc.13188>
- Golestani, N., Hervais-Adelman, A., Obleser, J., & Scott, S. K. (2013). Semantic versus perceptual interactions in neural processing of speech-in-noise. *NeuroImage*, 79, 52-61. <https://doi.org/10.1016/j.neuroimage.2013.04.049>
- Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience*.
- Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-Generality of Timing-Based Serial Order Processes in Short-Term Memory : New Insights from Musical and Verbal Domains. *PLOS ONE*, 11(12), e0168699. <https://doi.org/10.1371/journal.pone.0168699>
- Gorin, S., Mengal, P., & Majerus, S. (2018). Temporal grouping effects in musical short-term memory. *Memory*, 26(6), 831-843. <https://doi.org/10.1080/09658211.2017.1414848>
- Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired Memory for Pitch in Congenital Amusia. *Annals of the New York Academy of Sciences*, 1169(1), 270-272. <https://doi.org/10.1111/j.1749-6632.2009.04762.x>
- Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound?: *NeuroReport*, 10(18), 3825-3830. <https://doi.org/10.1097/00001756-199912160-00019>
- Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., Zatorre, R. J., Robitaille, N., & Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch : Evidence from magnetoencephalography. *NeuroImage*, 94, 96-106. <https://doi.org/10.1016/j.neuroimage.2014.03.020>
- Gruber, O., & Von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited. *NeuroImage*, 19(3), 797-809. [https://doi.org/10.1016/S1053-8119\(03\)00089-2](https://doi.org/10.1016/S1053-8119(03)00089-2)
- Guenot, M., Isnard, J., Ryvlin, P., Fischer, C., Ostrowsky, K., Mauguiere, F., & Sindou, M. (2001). Neurophysiological monitoring for epilepsy surgery : The Talairach SEEG method. *Stereotactic and functional neurosurgery*, 77(1-4), 29-32. <https://doi.org/10.1159/000064595>
- Hale, S., Bronik, M. D., & Fry, A. F. (1997). Verbal and spatial working memory in school-age children : Developmental differences in susceptibility to interference. *Developmental Psychology*, 33(2), 364-371. <https://doi.org/10.1037/0012-1649.33.2.364>
- Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children: *Ear and Hearing*, 23(2), 159-165. <https://doi.org/10.1097/00003446-200204000-00008>

References

- Han, Y. M. Y., Chan, M.-C., Chan, M. M. Y., Yeung, M. K., & Chan, A. S. (2022). Effects of working memory load on frontal connectivity in children with autism spectrum disorder : A fNIRS study. *Scientific Reports*, 12(1), 1522. <https://doi.org/10.1038/s41598-022-05432-3>
- Hancock, A. S., Warren, C. M., Barrett, T. S., Bolton, D. A. E., & Gillam, R. B. (2023). Functional near-infrared spectroscopy measures of neural activity in children with and without developmental language disorder during a working memory task. *Brain and Behavior*, 13(2). <https://doi.org/10.1002/brb3.2895>
- Henry, L. A. (1994). The Relationship between Speech Rate and Memory Span in Children. *International Journal of Behavioral Development*, 17(1), 37-56. <https://doi.org/10.1177/016502549401700103>
- Henson, R., & Friston, K. (2007). Convolution Models for fMRI. In *Statistical Parametric Mapping* (p. 178-192). Elsevier. <https://doi.org/10.1016/B978-012372560-8/50014-0>
- Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity : Behavior, Function, and Structure. *Neuron*, 76(3), 486-502. <https://doi.org/10.1016/j.neuron.2012.10.011>
- Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, 27(7-8), 1145-1166. <https://doi.org/10.1080/01690965.2012.662280>
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of cognitive neuroscience*, 15(5), 673-682.
- Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory : Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7(4), 347-362. <https://doi.org/10.1111/j.2044-835X.1989.tb00811.x>
- Hocke, L., Oni, I., Duszynski, C., Corrigan, A., Frederick, B., & Dunn, J. (2018). Automated Processing of fNIRS Data—A Visual Guide to the Pitfalls and Consequences. *Algorithms*, 11(5), 67. <https://doi.org/10.3390/a11050067>
- Hoefl, F., Ueno, T., Reiss, A. L., Meyler, A., Whitfield-Gabrieli, S., Glover, G. H., Keller, T. A., Kobayashi, N., Mazaika, P., Jo, B., Just, M. A., & Gabrieli, J. D. E. (2007). Prediction of children’s reading skills using behavioral, functional, and structural neuroimaging measures. *Behavioral Neuroscience*, 121(3), 602-613. <https://doi.org/10.1037/0735-7044.121.3.602>
- Hoen, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., & Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech*

References

- Communication, 49(12), 905-916.
<https://doi.org/10.1016/j.specom.2007.05.008>
- Holcomb, H. (1998). Cerebral blood flow relationships associated with a difficult tone recognition task in trained normal volunteers. *Cerebral Cortex*, 8(6), 534-542. <https://doi.org/10.1093/cercor/8.6.534>
- Hoshi, Y. (2016). Hemodynamic signals in fNIRS. In *Progress in Brain Research* (Vol. 225, p. 153-179). Elsevier. <https://doi.org/10.1016/bs.pbr.2016.03.004>
- Hoshi, Y., & Tamura, M. (1993a). Detection of dynamic changes in cerebral oxygenation coupled to neuronal function during mental work in man. *Neuroscience Letters*, 150(1), 5-8. [https://doi.org/10.1016/0304-3940\(93\)90094-2](https://doi.org/10.1016/0304-3940(93)90094-2)
- Hoshi, Y., & Tamura, M. (1993b). Dynamic multichannel near-infrared optical imaging of human brain activity. *Journal of Applied Physiology*, 75(4), 1842-1846. <https://doi.org/10.1152/jappl.1993.75.4.1842>
- Howes, D. (1957). On the relation between the probability of a word as an association and in general linguistic usage. *The Journal of Abnormal and Social Psychology*, 54(1), 75-85. <https://doi.org/10.1037/h0043830>
- Hoyer, R. S., Elshafei, H., Hemmerlin, J., Bouet, R., & Bidet-Caulet, A. (2021). Why Are Children So Distractible? Development of Attention and Motor Control From Childhood to Adulthood. *Child Development*, 92(4). <https://doi.org/10.1111/cdev.13561>
- Hu, X.-S., Arredondo, M. M., Gomba, M., Confer, N., DaSilva, A. F., Johnson, T. D., Shalinsky, M., & Kovelman, I. (2015). Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children. *Journal of Biomedical Optics*, 20(12), 126003. <https://doi.org/10.1117/1.JBO.20.12.126003>
- Hukin, R. W., & Darwin, C. J. (1995). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception & Psychophysics*, 57(2), 191-196. <https://doi.org/10.3758/BF03206505>
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30(6), 685-701. [https://doi.org/10.1016/0749-596X\(91\)90032-F](https://doi.org/10.1016/0749-596X(91)90032-F)
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1217-1232. <https://doi.org/10.1037/0278-7393.23.5.1217>
- Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared

References

- spectroscopy. *Neurophotonics*, 3(1), 010401.
<https://doi.org/10.1117/1.NPh.3.1.010401>
- Hyde, K. L. (2006). Morphometry of the amusic brain : A two-site study. *Brain*, 129(10), 2562-2570. <https://doi.org/10.1093/brain/awl204>
- Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical Thickness in Congenital Amusia : When Less Is Better Than More. *The Journal of Neuroscience*, 27(47), 13028-13032.
<https://doi.org/10.1523/JNEUROSCI.3039-07.2007>
- Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI Evidence of an Abnormal Neural Network for Pitch Processing in Congenital Amusia. *Cerebral Cortex*, 21(2), 292-299. <https://doi.org/10.1093/cercor/bhq094>
- Ichikawa, H., Nakato, E., Kanazawa, S., Shimamura, K., Sakuta, Y., Sakuta, R., Yamaguchi, M. K., & Kakigi, R. (2014). Hemodynamic response of children with attention-deficit and hyperactive disorder (ADHD) to emotional facial expressions. *Neuropsychologia*, 63, 51-58.
<https://doi.org/10.1016/j.neuropsychologia.2014.08.010>
- Inoue, Y., Sakihara, K., Gunji, A., Ozawa, H., Kimiya, S., Shinoda, H., Kaga, M., & Inagaki, M. (2012). Reduced prefrontal hemodynamic response in children with ADHD during the Go/NoGo task : A NIRS study. *NeuroReport*, 23(2), 55-60.
<https://doi.org/10.1097/WNR.0b013e32834e664c>
- Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span : A normative study. *British Journal of Developmental Psychology*, 7(4), 377-380. <https://doi.org/10.1111/j.2044-835X.1989.tb00814.x>
- James, W. (1890). *The principles of psychology*, Vol I. Henry Holt and Co.
<https://doi.org/10.1037/10538-000>
- Jasińska, K. K., Berens, M. S., Kovelman, I., & Petitto, L. A. (2017). Bilingualism yields language-specific plasticity in left hemisphere's circuitry for learning to read in young children. *Neuropsychologia*, 98, 34-45.
<https://doi.org/10.1016/j.neuropsychologia.2016.11.018>
- Jasińska, K. K., & Petitto, L. A. (2013). How age of bilingual exposure can change the neural systems for language in the developing brain : A functional near infrared spectroscopy investigation of syntactic processing in monolingual and bilingual children. *Developmental Cognitive Neuroscience*, 6, 87-101. <https://doi.org/10.1016/j.dcn.2013.06.005>
- Jasińska, K. K., & Petitto, L. A. (2014). Development of Neural Systems for Reading in the Monolingual and Bilingual Brain : New Insights From Functional Near Infrared Spectroscopy Neuroimaging. *Developmental Neuropsychology*, 39(6), 421-439.
<https://doi.org/10.1080/87565641.2014.939180>

References

- Jasińska, K. K., Shuai, L., Lau, A. N. L., Frost, S., Landi, N., & Pugh, K. R. (2021). Functional connectivity in the developing language network in 4-year-old children predicts future reading ability. *Developmental Science*, 24(2). <https://doi.org/10.1111/desc.13041>
- Jöbsis, F. F. (1977). Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters. *Science*, 198(4323), 1264-1267. <https://doi.org/10.1126/science.929199>
- Johnston, R. S., Johnson, C., & Gray, C. (1987). The emergence of the word length effect in young children : The effects of overt and covert rehearsal. *British Journal of Developmental Psychology*, 5(3), 243-248. <https://doi.org/10.1111/j.2044-835X.1987.tb01059.x>
- Jones, G., Justice, L. V., Cabiddu, F., Lee, B. J., Iao, L.-S., Harrison, N., & Macken, B. (2020). Does short-term memory develop? *Cognition*, 198, 104200. <https://doi.org/10.1016/j.cognition.2020.104200>
- Jones, G., & Macken, B. (2018). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition*, 46(2), 216-229. <https://doi.org/10.3758/s13421-017-0759-3>
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and Frequency in Specific Language Impairment : Accuracy and Error Data from Two Nonword Repetition Tests. *Journal of Speech, Language, and Hearing Research*, 53(6), 1642-1655. [https://doi.org/10.1044/1092-4388\(2010/09-0222\)](https://doi.org/10.1044/1092-4388(2010/09-0222))
- Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal Working Memory Load Affects Regional Brain Activation as Measured by PET. *Journal of Cognitive Neuroscience*, 9(4), 462-475. <https://doi.org/10.1162/jocn.1997.9.4.462>
- Jourdan Moser, S., Cutini, S., Weber, P., & Schroeter, M. L. (2009). Right prefrontal brain activation due to Stroop interference is altered in attention-deficit hyperactivity disorder—A functional near-infrared spectroscopy study. *Psychiatry Research: Neuroimaging*, 173(3), 190-195. <https://doi.org/10.1016/j.psychresns.2008.10.003>
- Kaas, J. H., & Hackett, T. A. (1999). « What » and « where » processing in auditory cortex. *Nature Neuroscience*, 2(12), 1045-1047. <https://doi.org/10.1038/15967>
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.
- Kato, T., Kamei, A., Takashima, S., & Ozaki, T. (1993). Human Visual Cortical Function during Photic Stimulation Monitoring by Means of near-Infrared Spectroscopy. *Journal of Cerebral Blood Flow & Metabolism*, 13(3), 516-520. <https://doi.org/10.1038/jcbfm.1993.66>
- Kawakubo, Y., Kono, T., Takizawa, R., Kuwabara, H., Ishii-Takahashi, A., & Kasai, K. (2011). Developmental Changes of Prefrontal Activation in

References

- Humans : A Near-Infrared Spectroscopy Study of Preschool Children and Adults. *PLoS ONE*, 6(10), e25944.
<https://doi.org/10.1371/journal.pone.0025944>
- Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 20160101. <https://doi.org/10.1098/rstb.2016.0101>
- Keller, T. A., & Cowan, N. (1994). Developmental Increase in the Duration of Memory for Tone Pitch. *Developmental Psychology*, 30(6), 855-863.
- Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 635-645. <https://doi.org/10.1037/0278-7393.21.3.635>
- Kidd, G., & Colburn, H. S. (2017). Informational Masking in Speech Recognition. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Éds.), *The Auditory System at the Cocktail Party* (Vol. 60, p. 75-109). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_4
- Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational Masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Éds.), *Auditory Perception of Sound Sources* (Vol. 29, p. 143-189). Springer US. https://doi.org/10.1007/978-0-387-71305-2_6
- Köchel, A., Schöngaßner, F., Feierl-Gsodam, S., & Schienle, A. (2015). Processing of affective prosody in boys suffering from attention deficit hyperactivity disorder : A near-infrared spectroscopy study. *Social Neuroscience*, 10(6), 583-591. <https://doi.org/10.1080/17470919.2015.1017111>
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory : An fMRI study. *Human Brain Mapping*, 30(3), 859-873.
<https://doi.org/10.1002/hbm.20550>
- Kohno, S., Miyai, I., Seiyama, A., Oda, I., Ishikawa, A., Tsuneishi, S., Amita, T., & Shimizu, K. (2007). Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis. *Journal of Biomedical Optics*, 12(6), 062111.
<https://doi.org/10.1117/1.2814249>
- Kolk, S. M., & Rakic, P. (2022). Development of prefrontal cortex. *Neuropsychopharmacology*, 47(1), 41-57. <https://doi.org/10.1038/s41386-021-01137-9>
- Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical Effects on Spoken-Word Recognition in Children with Normal Hearing. *Ear & Hearing*, 31(1), 102-114. <https://doi.org/10.1097/AUD.0b013e3181b7892f>

References

- Kuśmierk, P., & Rauschecker, J. P. (2009). Functional Specialization of Medial Auditory Belt Cortex in the Alert Rhesus Monkey. *Journal of Neurophysiology*, 102(3), 1606-1622. <https://doi.org/10.1152/jn.00167.2009>
- Langel, J., Hakun, J., Zhu, D. C., & Ravizza, S. M. (2014). Functional specialization of the left ventral parietal cortex in working memory. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00440>
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds : Effects of acoustic features and auditory object category. *Journal of Neuroscience*, 30(22), 7604-7612. <https://doi.org/10.1523/JNEUROSCI.0296-10.2010>
- Leibold, L. J., & Buss, E. (2013). Children's Identification of Consonants in a Speech-Shaped Noise or a Two-Talker Masker. *Journal of Speech, Language, and Hearing Research*, 56(4), 1144-1155. [https://doi.org/10.1044/1092-4388\(2012/12-0011\)](https://doi.org/10.1044/1092-4388(2012/12-0011))
- Leibold, L. J., & Buss, E. (2019). Masked Speech Recognition in School-Age Children. *Frontiers in Psychology*, 10, 1981. <https://doi.org/10.3389/fpsyg.2019.01981>
- Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental Effects in Masking Release for Speech-in-Speech Perception Due to a Target/Masker Sex Mismatch. *Ear & Hearing*, 39(5), 935-945. <https://doi.org/10.1097/AUD.0000000000000554>
- Lévêque, Y., Lalitte, P., Fornoni, L., Pralus, A., Albouy, P., Bouchet, P., Caclin, A., & Tillmann, B. (2022). Tonal structures benefit short-term memory for real music : Evidence from non-musicians and individuals with congenital amusia. *Brain and Cognition*, 161, 105881. <https://doi.org/10.1016/j.bandc.2022.105881>
- Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, 24(1), 61-79. https://doi.org/10.1162/jocn_a_00140
- Linke, A. C., & Cusack, R. (2015). Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds. *Journal of Cognitive Neuroscience*, 27(7), 1322-1333. https://doi.org/10.1162/jocn_a_00780
- Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences*, 108(31), 12961-12966. <https://doi.org/10.1073/pnas.1102118108>

References

- Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America*, 117(5), 3091-3099. <https://doi.org/10.1121/1.1873913>
- Litovsky, R. Y. (2015). Development of the auditory system. In *Handbook of Clinical Neurology* (Vol. 129, p. 55-72). Elsevier. <https://doi.org/10.1016/B978-0-444-62630-1.00003-2>
- Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain : The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, 34(3), 269-284. <https://doi.org/10.1016/j.neubiorev.2009.07.008>
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone Deafness : A New Disconnection Syndrome? *The Journal of Neuroscience*, 29(33), 10215-10220. <https://doi.org/10.1523/JNEUROSCI.1701-09.2009>
- Lozito, J. P., & Mulligan, N. W. (2006). Exploring the role of attention during memory retrieval : Effects of semantic encoding and divided attention. *Memory & Cognition*, 34(5), 986-998. <https://doi.org/10.3758/BF03193246>
- Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, 8(02). <https://doi.org/10.1117/1.NPh.8.2.025008>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : A user's guide*, 2nd ed. (p. xix, 492). Lawrence Erlbaum Associates Publishers.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.-P., Pittet, A., & Clarke, S. (2001). Distinct Pathways Involved in Sound Recognition and Localization : A Human fMRI Study. *NeuroImage*, 14(4), 802-816. <https://doi.org/10.1006/nimg.2001.0888>
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia : The Importance of Serial Order. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01522>
- Majerus, S., Martinez Perez, T., & Oberauer, K. (2012). Two distinct origins of long-term learning effects in verbal short-term memory. *Journal of Memory and Language*, 66(1), 38-51. <https://doi.org/10.1016/j.jml.2011.07.006>
- Massoudi, R., Van Wanrooij, M. M., Versnel, H., & Van Opstal, A. J. (2015). Spectrotemporal response properties of core auditory cortex neurons in awake monkey. *PLoS One*, 10(2), e0116118. <https://doi.org/doi.org/10.1371/journal.pone.0116118>
- McKeown, D., & Wellsted, D. (2009). Auditory memory for timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 855-875. <https://doi.org/10.1037/a0013708>

References

- Meek, J. H., Firbank, M., Elwell, C. E., Atkinson, J., Braddick, O., & Wyatt, J. S. (1998). Regional Hemodynamic Responses to Visual Stimulation in Awake Infants. *Pediatric Research*, 43(6), 840-843.
<https://doi.org/10.1203/00006450-199806000-00019>
- Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology*, 47(9), 546-553.
<https://doi.org/10.1080/14992020802252261>
- Mercer, T., & McKeown, D. (2010). Updating and feature overwriting in short-term memory for timbre. *Attention, Perception, & Psychophysics*, 72(8), 2289-2303. <https://doi.org/10.3758/BF03196702>
- Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., & Leseman, P. P. M. (2015). Growth of verbal short-term memory of nonwords varying in phonotactic probability : A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language*, 84, 24-36.
<https://doi.org/10.1016/j.jml.2015.05.001>
- Meyer, T. A., & Pisoni, D. B. (1999). Some Computational Analyses of the PBK Test : Effects of Frequency and Lexical Density on Spoken Word Recognition: *Ear and Hearing*, 20(4), 363-371.
<https://doi.org/10.1097/00003446-199908000-00008>
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the Hemodynamic Response : Effects of Presentation Rate, Sampling Procedure, and the Possibility of Ordering Brain Activity Based on Relative Timing. *NeuroImage*, 11(6), 735-759.
<https://doi.org/10.1006/nimg.2000.0568>
- Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, 44(2), 105-129. <https://doi.org/10.1037/h0055960>
- Milner, B. (1966). Amnesia following operation on the temporal lobes. In *Amnesia*. C.W.M. Whitty, and O.L. Zangwill, eds. (London, Butterworth).
- Misurelli, S. M., & Litovsky, R. Y. (2015). Spatial release from masking in children with bilateral cochlear implants and with normal hearing : Effect of target-interferer similarity. *The Journal of the Acoustical Society of America*, 138(1), 319-331. <https://doi.org/10.1121/1.4922777>
- Monden, Y., Dan, I., Nagashima, M., Dan, H., Uga, M., Ikeda, T., Tsuzuki, D., Kyutoku, Y., Gunji, Y., Hirano, D., Taniguchi, T., Shimoizumi, H., Watanabe, E., & Yamagata, T. (2015). Individual classification of ADHD children by right prefrontal hemodynamic responses during a go/no-go task as assessed by fNIRS. *NeuroImage: Clinical*, 9, 1-12.
<https://doi.org/10.1016/j.nicl.2015.06.011>
- Moore, D. R., Cowan, J. A., Riley, A., Edmondson-Jones, A. M., & Ferguson, M. A. (2011). Development of Auditory Processing in 6- to 11-Yr-Old

References

- Children. *Ear & Hearing*, 32(3), 269-285.
<https://doi.org/10.1097/AUD.0b013e318201c468>
- Moray, N. (1960). Beoadbent's Filter Theory : Postulate H and the Problem of Switching Time. *Quarterly Journal of Experimental Psychology*, 12(4), 214-220. <https://doi.org/10.1080/17470216008416728>
- Mücke, M., Andrä, C., Gerber, M., Pühse, U., & Ludyga, S. (2018). Moderate-to-vigorous physical activity, executive functions and prefrontal brain oxygenation in children : A functional near-infrared spectroscopy study. *Journal of Sports Sciences*, 36(6), 630-636.
<https://doi.org/10.1080/02640414.2017.1326619>
- Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall : The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 594-598. <https://doi.org/10.1037/0278-7393.28.3.594>
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology : Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science*, 31(6), 678-701. <https://doi.org/10.1177/0956797620916782>
- Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E. (2005). The Role of the Prefrontal Cortex in the Maintenance of Verbal Working Memory : An Event-Related fMRI Analysis. *Neuropsychology*, 19(2), 223-232. <https://doi.org/10.1037/0894-4105.19.2.223>
- Nippert, A. R., Biesecker, K. R., & Newman, E. A. (2018). Mechanisms Mediating Functional Hyperemia in the Brain. *The Neuroscientist*, 24(1), 73-83. <https://doi.org/10.1177/1073858417703033>
- Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading Disabilities in SLI and Dyslexia Result From Distinct Phonological Impairments. *Developmental Neuropsychology*, 34(3), 296-311. <https://doi.org/10.1080/87565640902801841>
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm : A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46-59.
<https://doi.org/10.1002/hbm.20131>
- Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6), 1185-1191. <https://doi.org/10.1037/0278-7393.15.6.1185>

References

- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342-345. <https://doi.org/10.1038/362342a0>
- Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch : Implications for a working-memory model. *Memory & Cognition*, 20(3), 314-320. <https://doi.org/10.3758/BF03199668>
- Peretz, I. (2016). Neurobiology of Congenital Amusia. *Trends in Cognitive Sciences*, 20(11), 857-867. <https://doi.org/10.1016/j.tics.2016.09.002>
- Peretz, I., & Zatorre, R. J. (2005). Brain Organization for Music Processing. *Annual Review of Psychology*, 56(1), 89-114. <https://doi.org/10.1146/annurev.psych.56.091103.070225>
- Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage*, 62(2), 1177-1184. <https://doi.org/10.1016/j.neuroimage.2011.09.084>
- Pickering, S. J., Gathercole, S. E., Hall, M., & Lloyd, S. A. (2001). Development of Memory for Pattern and Path : Further Evidence for the Fractionation of Visuo-Spatial Memory. *The Quarterly Journal of Experimental Psychology Section A*, 54(2), 397-420. <https://doi.org/10.1080/713755973>
- Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data : An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, 12, 505. <https://doi.org/10.3389/fnhum.2018.00505>
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1), 5-29. <https://doi.org/10.1111/nyas.13948>
- Pisoni, A., Mattavelli, G., Casarotti, A., Comi, A., Riva, M., Bello, L., & Papagno, C. (2019). The neural correlates of auditory-verbal short-term memory : A voxel-based lesion-symptom mapping study on 103 patients after glioma removal. *Brain Structure and Function*, 224(6), 2199-2211. <https://doi.org/10.1007/s00429-019-01902-z>
- Plantinga, J., & Trainor, L. J. (2008). Infants' Memory for Isolated Tones and the Effects of Interference. *Music Perception: An Interdisciplinary Journal*, 26(2), 121-127. <https://doi.org/10.1525/mp.2008.26.2.121>
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows : Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41(1), 245-255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)

References

- Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, 57(S1), S5-S5. <https://doi.org/10.1121/1.1995329>
- Pollonini, L., Bortfeld, H., & Oghalai, J. S. (2016). PHOEBE: A method for real time mapping of optodes-scalp coupling in functional near-infrared spectroscopy. *Biomedical Optics Express*, 7(12), 5104. <https://doi.org/10.1364/BOE.7.005104>
- Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., & Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hearing Research*, 309, 84-93. <https://doi.org/10.1016/j.heares.2013.11.007>
- Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., & Mishkin, M. (2003). Functional Mapping of the Primate Auditory System. *Science*, 299(5606), 568-572. <https://doi.org/10.1126/science.1078900>
- Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816-847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>
- Pujol, J., Soriano-Mas, C., Ortiz, H., Sebastian-Galles, N., Losilla, J. M., & Deus, J. (2006). Myelination of language-related areas in the developing brain. *Neurology*, 66(3), 339-343. <https://doi.org/10.1212/01.wnl.0000201049.66073.8d>
- Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13(2), 212-218. [https://doi.org/10.1016/S0959-4388\(03\)00035-7](https://doi.org/10.1016/S0959-4388(03)00035-7)
- Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, 9(8), 374-380. <https://doi.org/10.1016/j.tics.2005.06.009>
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516-521. [https://doi.org/10.1016/S0959-4388\(98\)80040-8](https://doi.org/10.1016/S0959-4388(98)80040-8)
- Rauschecker, J. P. (2015). Auditory Cortex. In *Brain Mapping* (p. 299-304). Elsevier. <https://doi.org/10.1016/B978-0-12-397025-1.00226-8>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724. <https://doi.org/10.1038/nn.2331>
- Rauschecker, J. P., & Tian, B. (2004). Processing of Band-Passed Noise in the Lateral Auditory Belt Cortex of the Rhesus Monkey. *Journal of Neurophysiology*, 91(6), 2578-2589. <https://doi.org/10.1152/jn.00834.2003>
- Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., & Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal

References

- working memory. *NeuroImage*, 22(2), 562-573.
<https://doi.org/10.1016/j.neuroimage.2004.01.039>
- Ravizza, S. M., Hazeltine, E., Ruiz, S., & Zhu, D. C. (2011). Left TPJ activity in verbal working memory : Implications for storage- and sensory-specific models of short term memory. *NeuroImage*, 55(4), 1836-1846.
<https://doi.org/10.1016/j.neuroimage.2010.12.021>
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The Neural Mechanisms of Speech Comprehension : fMRI studies of Semantic Ambiguity. *Cerebral Cortex*, 15(8), 1261-1269. <https://doi.org/10.1093/cercor/bhi009>
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers : Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431-2443. <https://doi.org/10.1121/1.4794379>
- Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory : A coordinate-based meta-analysis. *NeuroImage*, 60(1), 830-846.
<https://doi.org/10.1016/j.neuroimage.2011.11.050>
- Russ, B. E., Ackelson, A. L., Baker, A. E., & Cohen, Y. E. (2008). Coding of Auditory-Stimulus Identity in the Auditory Non-Spatial Processing Stream. *Journal of Neurophysiology*, 99(1), 87-95.
<https://doi.org/10.1152/jn.01069.2007>
- Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory : Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences*, 96(11), 6558-6563. <https://doi.org/10.1073/pnas.96.11.6558>
- Rypma, B., Prabhakaran, V., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Load-Dependent Roles of Frontal Brain Regions in the Maintenance of Working Memory. *NeuroImage*, 9(2), 216-226.
<https://doi.org/10.1006/nimg.1998.0404>
- Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, 30(8), 1185-1194.
<https://doi.org/10.1080/00140138708966007>
- Salamé, P., & Baddeley, A. D. (1989). Effects of Background Music on Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology Section A*, 41(1), 107-122.
<https://doi.org/10.1080/14640748908402355>
- Salmon, E., Van Der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., Luxen, A., & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain*, 119(5), 1617-1625.
<https://doi.org/10.1093/brain/119.5.1617>

References

- Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms*, 11(5), 73.
- Santosa, H., Zhai, X., Fishburn, F., Sparto, P. J., & Huppert, T. J. (2020). Quantitative comparison of correction techniques for removing systemic physiological signal in functional near-infrared spectroscopy studies. *Neurophotonics*, 7(03). <https://doi.org/10.1117/1.NPh.7.3.035009>
- Sato, T., Nambu, I., Takeda, K., Aihara, T., Yamashita, O., Isogaya, Y., Inoue, Y., Otaka, Y., Wada, Y., Kawato, M., Sato, M., & Osu, R. (2016). Reduction of global interference of scalp-hemodynamics in functional near-infrared spectroscopy using short distance probes. *NeuroImage*, 141, 120-132. <https://doi.org/10.1016/j.neuroimage.2016.06.054>
- Savin, H. B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America*, 35(2), 200-206. <https://doi.org/10.1121/1.1918432>
- Schecklmann, M., Romanos, M., Bretscher, F., Plichta, M. M., Warnke, A., & Fallgatter, A. J. (2010). Prefrontal oxygenation during working memory in ADHD. *Journal of Psychiatric Research*, 44(10), 621-628. <https://doi.org/10.1016/j.jpsychires.2009.11.018>
- Schecklmann, M., Schaldecker, M., Aucktor, S., Brast, J., Kirchgäßner, K., Mühlberger, A., Warnke, A., Gerlach, M., Fallgatter, A. J., & Romanos, M. (2011). Effects of methylphenidate on olfaction and frontal and temporal brain oxygenation in children with ADHD. *Journal of Psychiatric Research*, 45(11), 1463-1470. <https://doi.org/10.1016/j.jpsychires.2011.05.011>
- Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, 35(4), 640-650. <https://doi.org/10.3758/BF03193302>
- Schlaug, G. (2001). The Brain of Musicians: A Model for Functional and Structural Adaptation. *Annals of the New York Academy of Sciences*, 930(1), 281-299. <https://doi.org/10.1111/j.1749-6632.2001.tb05739.x>
- Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. In *Progress in Brain Research* (Vol. 217, p. 37-55). Elsevier. <https://doi.org/10.1016/bs.pbr.2014.11.020>
- Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, 85, 6-27. <https://doi.org/10.1016/j.neuroimage.2013.05.004>
- Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, 106(34), 14611-14616. <https://doi.org/10.1073/pnas.0907682106>

References

- Schroeter, M. L., Bücheler, M. M., Müller, K., Uludağ, K., Obrig, H., Lohmann, G., Tittgemeyer, M., Villringer, A., & Von Cramon, D. Y. (2004). Towards a standard analysis for functional near-infrared imaging. *NeuroImage*, 21(1), 283-290. <https://doi.org/10.1016/j.neuroimage.2003.09.054>
- Schulze, K., Jay Dowling, W., & Tillmann, B. (2012). Working Memory for Tonal and Atonal Sequences during a Forward and a Backward Recognition Task. *Music Perception*, 29(3), 255-267. <https://doi.org/10.1525/mp.2012.29.3.255>
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1), 229-236. <https://doi.org/10.1111/j.1749-6632.2012.06447.x>
- Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, 21(3), 377-395. <https://doi.org/10.1080/09658211.2012.731070>
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32(5), 771-783. <https://doi.org/10.1002/hbm.21060>
- Semal, C., Demany, L., Ueda, K., & Hallé, P.-A. (1996). Speech versus nonspeech in pitch memory. *The Journal of the Acoustical Society of America*, 100(2), 1132-1140. <https://doi.org/10.1121/1.416298>
- Shallice, T., & Papagno, C. (2019). Impairments of auditory-verbal short-term memory : Do selective deficits of the input phonological buffer exist? *Cortex*, 112, 107-121. <https://doi.org/10.1016/j.cortex.2018.10.004>
- Shallice, T., & Warrington, E. K. (1970). Independent Functioning of Verbal Memory Stores : A Neuropsychological Study. *Quarterly Journal of Experimental Psychology*, 22(2), 261-273. <https://doi.org/10.1080/00335557043000203>
- Sharp, D. J., Scott, S. K., Mehta, M. A., & Wise, R. J. S. (2005). The Neural Correlates of Declining Performance with Age : Evidence for Age-Related Changes in Cognitive Control. *Cerebral Cortex*, 16(12), 1739-1749. <https://doi.org/10.1093/cercor/bhj109>
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, 118(5), 2775-2778. <https://doi.org/10.1121/1.2062650>
- Soltanlou, M., Sitnikova, M. A., Nuerk, H.-C., & Dresler, T. (2018). Applications of Functional Near-Infrared Spectroscopy (fNIRS) in Studying Cognitive Development : The Case of Mathematics and Language. *Frontiers in Psychology*, 9, 277. <https://doi.org/10.3389/fpsyg.2018.00277>

References

- Song, R., Zhang, J., Wang, B., Zhang, H., & Wu, H. (2013). A near-infrared brain function study of Chinese dyslexic children. *Neurocase*, 19(4), 382-389. <https://doi.org/10.1080/13554794.2012.690422>
- Stewart, L. (2008). Do musicians have different brains? *Clinical Medicine*, 8(3), 304-308. <https://doi.org/10.7861/clinmedicine.8-3-304>
- Strangman, G. E., Li, Z., & Zhang, Q. (2013). Depth Sensitivity and Source-Detector Separations for Near Infrared Spectroscopy Based on the Colin27 Brain Template. *PLoS ONE*, 8(8), e66319. <https://doi.org/10.1371/journal.pone.0066319>
- Sugiura, L., Ojima, S., Matsuba-Kurita, H., Dan, I., Tsuzuki, D., Katura, T., & Hagiwara, H. (2011). Sound to Language: Different Cortical Processing for First and Second Languages in Elementary School Children as Revealed by a Large-Scale Study Using fNIRS. *Cerebral Cortex*, 21(10), 2374-2393. <https://doi.org/10.1093/cercor/bhr023>
- Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *Journal of Speech, Language, and Hearing Research*, 60(10), 2989-3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041
- Tak, S., & Ye, J. C. (2014). Statistical analysis of fNIRS data: A comprehensive review. *NeuroImage*, 85, 72-91. <https://doi.org/10.1016/j.neuroimage.2013.06.016>
- Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: Influence of material type, contour, and musical expertise. *Psychological Research*. <https://doi.org/10.1007/s00426-021-01519-0>
- Tando, T., Kaga, Y., Ishii, S., Aoyagi, K., Sano, F., Kanemura, H., Sugita, K., & Aihara, M. (2014). Developmental changes in frontal lobe function during a verbal fluency task: A multi-channel near-infrared spectroscopy study. *Brain and Development*, 36(10), 844-852. <https://doi.org/10.1016/j.braindev.2014.01.002>
- Tellis, G., & Tellis, C. (2016). Using Functional near Infrared Spectroscopy with Fluent Speakers to Determine Haemoglobin Changes in the Brain during Speech and Non-Speech Tasks. *NIR News*, 27(3), 4-7. <https://doi.org/10.1255/nirn.1599>
- Tian, B., & Rauschecker, J. P. (2004). Processing of Frequency-Modulated Sounds in the Lateral Auditory Belt Cortex of the Rhesus Monkey. *Journal of Neurophysiology*, 92(5), 2993-3013. <https://doi.org/10.1152/jn.00472.2003>
- Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. In *Handbook of Clinical Neurology* (Vol. 129, p. 589-605). Elsevier. <https://doi.org/10.1016/B978-0-444-62630-1.00033-0>

References

- Tillmann, B., Graves, J. E., Talamini, F., Leveque, Y., Fornoni, L., Hoarau, C., Pralus, A., Ginzburg, J., Albouy, P., & Caclin, A. (2023). Auditory cortex and beyond: Deficits in congenital amusia. *Hearing Research*, 437, 108855. <https://doi.org/10.1016/j.heares.2023.108855>
- Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, 1640, 251-263. <https://doi.org/10.1016/j.brainres.2015.10.035>
- Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, 71(3), 259-264. <https://doi.org/10.1016/j.bandc.2009.08.003>
- Treisman, A. M. (1960). Contextual Cues in Selective Listening. *Quarterly Journal of Experimental Psychology*, 12(4), 242-248. <https://doi.org/10.1080/17470216008416732>
- Tsujimoto, S., Yasumura, A., Yamashita, Y., Torii, M., Kaga, M., & Inagaki, M. (2013). Increased Prefrontal Oxygenation Related to Distractor-Resistant Working Memory in Children with Attention-Deficit/Hyperactivity Disorder (ADHD). *Child Psychiatry & Human Development*, 44(5), 678-688. <https://doi.org/10.1007/s10578-013-0361-2>
- Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The Cingulo-Opercular Network Provides Word-Recognition Benefit. *The Journal of Neuroscience*, 33(48), 18979-18986. <https://doi.org/10.1523/JNEUROSCI.1417-13.2013>
- Van Viersen, S., De Bree, E. H., & De Jong, P. F. (2019). Protective Factors and Compensation in Resolving Dyslexia. *Scientific Studies of Reading*, 23(6), 461-477. <https://doi.org/10.1080/10888438.2019.1603543>
- Vickers, D. A., Moore, B. C. J., Majeed, A., Stephenson, N., Alferaih, H., Baer, T., & Marriage, J. E. (2018). Closed-Set Speech Discrimination Tests for Assessing Young Children. *Ear & Hearing*, 39(1), 32-41. <https://doi.org/10.1097/AUD.0000000000000528>
- Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, 20(10), 435-442. [https://doi.org/10.1016/S0166-2236\(97\)01132-6](https://doi.org/10.1016/S0166-2236(97)01132-6)
- Villringer, A., Planck, J., Stodieck, S., Bötzel, K., Schleinkofer, L., & Dirnagl, U. (1994). Noninvasive Assessment of Cerebral Hemodynamics and Tissue Oxygenation during Activation of Brain Cell Function in Human Adults Using Near Infrared Spectroscopy. In P. Vaupel, R. Zander, & D. F. Bruley (Éds.), *Oxygen Transport to Tissue XV* (Vol. 345, p. 559-565). Springer US. https://doi.org/10.1007/978-1-4615-2468-7_74
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall.

References

- Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5), 1256-1271. <https://doi.org/10.1037/0278-7393.25.5.1256>
- Wang, J., Zhang, C., Wan, S., & Peng, G. (2017). Is Congenital Amusia a Disconnection Syndrome? A Study Combining Tract- and Network-Based Analysis. *Frontiers in Human Neuroscience*, 11, 473. <https://doi.org/10.3389/fnhum.2017.00473>
- Wang, X. (2000). On cortical coding of vocal communication sounds in primates. *Proceedings of the National Academy of Sciences*, 97(22), 11843-11849. <https://doi.org/10.1073/pnas.97.22.11843>
- Webb, E. K., Etter, J. A., & Kwasa, J. A. (2022). Addressing racial and phenotypic bias in human neuroscience methods. *Nature Neuroscience*, 25(4), 410-414. <https://doi.org/10.1038/s41593-022-01046-0>
- Wegel, R. L., & Lane, C. E. (1924). The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Physical Review*, 23(2), 266-285. <https://doi.org/10.1103/PhysRev.23.266>
- Weiss, A. H., Granot, R. Y., & Ahissar, M. (2014). The enigma of dyslexic musicians. *Neuropsychologia*, 54, 28-40. <https://doi.org/10.1016/j.neuropsychologia.2013.12.009>
- Werner, L., Fay, R. R., & Popper, A. N. (Éds.). (2012). *Human Auditory Development* (Vol. 42). Springer New York. <https://doi.org/10.1007/978-1-4614-1421-6>
- Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children : Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, 118(5), 3164-3176. <https://doi.org/10.1121/1.2082567>
- Wilcox, T., & Biondi, M. (2015). fNIRS in the developmental sciences : fNIRS in the developmental sciences. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(3), 263-283. <https://doi.org/10.1002/wcs.1343>
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful Listening : The Processing of Degraded Speech Depends Critically on Attention. *The Journal of Neuroscience*, 32(40), 14010-14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>
- Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences : Comparing phonological similarity and pitch proximity. *Memory & Cognition*, 38(2), 163-175. <https://doi.org/10.3758/MC.38.2.163>
- Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology*, 6(1), 15-22. <https://doi.org/10.2478/v10053-008-0073-5>

References

- Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia : Beyond a fine-grained pitch discrimination problem. *Memory*, 18(6), 657-669. <https://doi.org/10.1080/09658211.2010.501339>
- Wilson, W. J., & Arnott, W. (2013). Using different criteria to diagnose (central) auditory processing disorder : How big a difference does it make? *Journal of Speech, Language, and Hearing Research*, 56, 63-70. [https://doi.org/10.1044/1092-4388\(2012/11-0352\)](https://doi.org/10.1044/1092-4388(2012/11-0352))
- Wong, P. C. M., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical Mechanisms of Speech Perception in Noise. *Journal of Speech, Language, and Hearing Research*, 51(4), 1026-1041. [https://doi.org/10.1044/1092-4388\(2008/075\)](https://doi.org/10.1044/1092-4388(2008/075))
- Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory : A role for (co-)articulatory fluency? *Journal of Memory and Language*, 58(1), 48-65. <https://doi.org/10.1016/j.jml.2007.07.002>
- Xiao, T., Xiao, Z., Ke, X., Hong, S., Yang, H., Su, Y., Chu, K., Xiao, X., Shen, J., & Liu, Y. (2012). Response Inhibition Impairment in High Functioning Autism and Attention Deficit Hyperactivity Disorder : Evidence from Near-Infrared Spectroscopy Data. *PLoS ONE*, 7(10), e46569. <https://doi.org/10.1371/journal.pone.0046569>
- Xu, S.-Y., Lu, F.-M., Wang, M.-Y., Hu, Z.-S., Zhang, J., Chen, Z.-Y., Armada-da-Silva, P. A. S., & Yuan, Z. (2020). Altered Functional Connectivity in the Motor and Prefrontal Cortex for Children With Down's Syndrome : An fNIRS Study. *Frontiers in Human Neuroscience*, 14, 6. <https://doi.org/10.3389/fnhum.2020.00006>
- Ye, J., Tak, S., Jang, K., Jung, J., & Jang, J. (2009). NIRS-SPM : Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*, 44(2), 428-447. <https://doi.org/10.1016/j.neuroimage.2008.08.036>
- Yeung, M. K., Lee, T. L., & Chan, A. S. (2019). Right-lateralized frontal activation underlies successful updating of verbal working memory in adolescents with high-functioning autism spectrum disorder. *Biological Psychology*, 148, 107743. <https://doi.org/10.1016/j.biopsycho.2019.107743>
- Yue, Q., Martin, R. C., Hamilton, A. C., & Rose, N. S. (2019). Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory : Evidence for a Buffer Account? *Cerebral Cortex*, 29(4), 1398-1413. <https://doi.org/10.1093/cercor/bhy037>
- Yuen, K. C. P., & Yuan, M. (2014). Development of Spatial Release From Masking in Mandarin-Speaking Children With Normal Hearing. *Journal of Speech, Language, and Hearing Research*, 57(5), 2005-2023. https://doi.org/10.1044/2014_JSLHR-H-13-0060

References

- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, 11(10), 946-953.
<https://doi.org/10.1093/cercor/11.10.946>
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex : Music and speech. *Trends in Cognitive Sciences*, 6(1), 37-46. [https://doi.org/10.1016/S1364-6613\(00\)01816-7](https://doi.org/10.1016/S1364-6613(00)01816-7)
- Zatorre, R. J., Evans, A., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4), 1908-1919. <https://doi.org/10.1523/JNEUROSCI.14-04-01908.1994>
- Zatorre, R. J., & Samson, S. (1991). ROLE OF THE RIGHT TEMPORAL NEOCORTEX IN RETENTION OF PITCH IN AUDITORY SHORT-TERM MEMORY. *Brain*, 114(6), 2403-2417.
<https://doi.org/10.1093/brain/114.6.2403>
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32(4), 1826-1836. <https://doi.org/10.1016/j.neuroimage.2006.04.199>
- Zhang, F., & Roeyers, H. (2019). Exploring brain functions in autism spectrum disorder : A systematic review on functional near-infrared spectroscopy (fNIRS) studies. *International Journal of Psychophysiology*, 137, 41-53.
<https://doi.org/10.1016/j.ijpsycho.2019.01.003>
- Zhang, Y., Brooks, D. H., Franceschini, M. A., & Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *Journal of Biomedical Optics*, 10(1), 011014. <https://doi.org/10.1117/1.1852552>
- Zhao, Y., Chen, X., Zhong, S., Cui, Z., Gong, G., Dong, Q., & Nan, Y. (2016). Abnormal topological organization of the white matter network in Mandarin speakers with congenital amusia. *Scientific Reports*, 6(1), 26505.
<https://doi.org/10.1038/srep26505>
- Zhu, H., Fan, Y., Guo, H., Huang, D., & He, S. (2014). Reduced interhemispheric functional connectivity of children with autism spectrum disorder : Evidence from functional near infrared spectroscopy studies. *Biomedical Optics Express*, 5(4), 1262. <https://doi.org/10.1364/BOE.5.001262>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Foxton, J. M. (2012). Global and local pitch perception in children with developmental dyslexia. *Brain and Language*, 120(3), 265-270. <https://doi.org/10.1016/j.bandl.2011.12.002>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12(5), 732-745. <https://doi.org/10.1111/j.1467-7687.2009.00817.x>
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off : Speech perception deficits in children with specific language

References

impairment. *Journal of Experimental Child Psychology*, 110(3), 362-372.
<https://doi.org/10.1016/j.jecp.2011.05.001>