HAL
open science

Non-verbal auditory cognition deficits : from assessment to the development of a new rehabilitation program

Agathe Pralus-Durand

To cite this version:

Agathe Pralus-Durand. Non-verbal auditory cognition deficits : from assessment to the development of a new rehabilitation program. Neurons and Cognition [q-bio.NC]. Université de Lyon, 2020. English. NNT : 2020LYSE1180 . tel-03349203

HAL Id: tel-03349203

https://theses.hal.science/tel-03349203v1

Submitted on 20 Sep 2021

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.

N° d'ordre NNT :

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N°476
(Neurosciences et cognition)

Spécialité de doctorat : Neurosciences

Soutenue publiquement le 08/10/2020, par :
(Agathe Audrey Danielle Pralus-Durand)

Non-verbal auditory cognition deficits: from assessment to the development of a new rehabilitation program

Devant le jury composé de :

Dr Pascal Belin (Rapporteur)
Dr Daniela Sammler (Rapporteure)
Dr Séverine Samson (Examinatrice)
Pr Eric Truy (Examinateur)
Dr Barbara Tillmann (Directrice de thèse)
Dr Anne Caclin (Co-directrice de thèse)

Université Claude Bernard – LYON 1

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

Remerciements

Mes premiers remerciements vont bien entendu à mes deux chères directrices de thèse, Anne et Barbara. Je ne vous remercierai jamais assez pour m'avoir fait confiance dès le début pour ce travail de thèse, pour avoir cru en mes idées de jeune recrue et pour m'avoir soutenue dans tous mes projets. J'ai tellement appris à vos côtés que ce soit au niveau technique mais aussi sur moi-même et mes capacités. Tous vos étudiants et futurs étudiants ont beaucoup de chance de vous avoir à leurs côtés. Merci d'avoir apporté votre soutien moral et compréhension, même quand des petites surprises sont arrivées en cours de thèse !

Je remercie chaleureusement Daniel Sammler et Pascal Belin pour avoir accepté d'être rapporteurs de ce travail, ainsi que Séverine Samson et Eric Truy pour leur participation au jury.

Je tiens à remercier l'HEH ainsi que tout son personnel pour avoir accueilli notre projet avec bienveillance et avoir permis sa réalisation, merci à Ruben Hermann et Eric Truy. Bien évidemment, un grand merci à l'équipe DYCOG et plus particulièrement à mes camarades de la PAMTeam ! Tous les pampam girls and boys m'ont donné chaque jour envie de venir au bureau, voir vos sourires m'a donné le courage de tenir jusqu'au bout malgré les jours de découragement. Notre bureau était pour moi un havre de paix où l'on pouvait relâcher la pression, rire, mais aussi échanger des idées (plus ou moins) brillantes ! Merci à Salomé et Roxane pour leurs jolies créations en crochet qui nous ont accompagnés toute l'année sur nos bureaux, merci à Rémy pour ces cultures de champignon qui remplissaient l'atmosphère d'une odeur bien à lui et merci à Lou d'avoir apporté de la beauté et de la bonne humeur tous les jours (et on sait que ce n'était pas si simple !!). Merci à nos « mamans » PAM qui ont toujours veillé sur nous du début à la fin, vous êtes des vraies battantes et j'étais plus que fière de faire partie de notre équipe de filles !! Un merci tout particulier à Lesly, ma compagne de manips et de galère amusique, tu es le roc de cette équipe.

J'envoies beaucoup de remerciements et de câlins à ma famille, vous m'avez toujours soutenue depuis le début, même quand moi je n'y croyais plus. Merci d'avoir supporté mes bêtises neuroscientifiques, même sans comprendre vous savez toujours trouvé les bons mots pour relancer ma détermination. Merci Mamounette, Papou, Pau et Tessouille !

Enfin, comment ne pas remercier mon meilleur soutien tout au long de ce travail de longue haleine… Un merci sans fin à mon mari, mon partenaire de débat et ma motivation de toujours, Maxime. Tu as su me supporter pendant toutes ces années sans (presque) broncher, et tout ceci n'aurait pas été possible sans toi. Grâce à toi, j'ai eu la joie infinie d'accueillir la plus merveilleuse fille qui soit sans que mon travail n'en soit affecté. Tu m'as permis d'être une maman comblée et fière d'elle ! Merci à toi et merci à mon Aelys adorée pour ses petits rires qui ont toujours su me rappeler que la vie est magnifique.

Abstract

The aim of this PhD thesis is to further characterize non-verbal auditory cognition, in particular when it is impaired. The first axis of this research focused on the investigation of emotion perception in two populations with potential deficits: in individuals with congenital amusia and patients with brain-damage. In congenital amusia, Study 1 demonstrated a deficit of emotional prosody perception. This deficit was specifically present for short vowels (versus long sentences). However, the deficit was only present in the explicit recognition task and not in the implicit intensity ratings of the same emotions. Moreover, Study 2 allowed us to relate this explicit recognition deficit with early automatic brain processing decrease by using electroencephalography. In brain-damaged patients, Study 3 demonstrated a deficit for musical emotion perception in relation with the side of the lesion. Based on these results and previous studies, we decided to design a new rehabilitation strategy for the training of non-verbal auditory cognition. In the second axis of this thesis, we focused on developing a new training strategy and chose to test this new training with cochlear implant (CI) users as this population is in high demand for better auditory cognition. We first designed a new short assessment battery for non-verbal auditory cognition. Study 4 demonstrated its efficiency to reveal specific deficits in CI users and in normal-hearing participants by using vocoded sounds. Moreover, findings revealed some evidence that audiovisual cues might help CI users to enhance their non-verbal auditory perception, as previously suggested with verbal material. We then designed a new training strategy by using multisensory integration and more specifically audiovisual stimulation. We suggest that this training could enhance non-verbal auditory abilities of CI users, but also of control participants. We aim to demonstrate the efficiency of this training in a long-term implementation by acquiring both behavioural measures (with the assessment battery developed in Study 4) in all populations with a deficit, but also magnetoencephalographic measures in control participants (Study 5). Overall, this PhD research brings further insight in the field of non-verbal auditory cognition and its associated deficits, and provides a new tool aiming to measure and remediate these deficits, which will then be useable for evaluation in clinical settings.

Keywords: non-verbal auditory cognition, emotional prosody, musical emotion, pitch, implicit knowledge, multisensory integration, audiovisual, electroencephalography, training, rehabilitation, magnetoencephalography

Résumé

Cette thèse a pour but de fournir de nouveaux outils pour la caractérisation de la cognition auditive non-verbale, ses déficits, et développer un outil pour la réhabilitation. Nous avons choisi de concentrer le premier axe de ce travail sur une meilleure compréhension de la perception des émotions dans deux déficits : l'amusie congénitale et à la suite d'une lésion cérébrale. Dans l'amusie congénitale, nous avons démontré un déficit de perception de la prosodie émotionnelle, en utilisant de courtes voyelles plutôt que de longues phrases. Ce déficit de perception de la prosodie émotionnelle était présent uniquement dans la tâche de reconnaissances des émotions, les amusiques démontrant des notations d'intensité similaires aux contrôles. De plus, nous avons associé ce déficit à une diminution du traitement cérébral automatique précoce grâce à des mesures en électroencéphalographie. Chez les patients présentant une lésion cérébrale, nous avons démontré un déficit de perception des émotions musicales en relation avec le côté de la lésion. Sur la base de ces résultats et d'études antérieures, nous avons décidé de concevoir une nouvelle stratégie de rééducation pour la cognition auditive non verbale. Dans le deuxième axe de cette thèse, nous nous sommes concentrés sur les possibilités d'entrainement de cette cognition et avons choisi de tester cet entrainement avec des utilisateurs d'implants cochléaires (IC). En effet, au sein de cette population, de nombreuses personnes demandent des améliorations pour mieux percevoir les informations non verbales. Dans une première étude, nous avons conçu une nouvelle batterie d'évaluation de la cognition auditive non verbale. Nous avons démontré son efficacité pour révéler des déficits spécifiques chez les utilisateurs d'IC ainsi que chez les participants ayant une audition normale en utilisant des sons vocodés. De plus, nous avons apporté des preuves que les indices audiovisuels pourraient aider les utilisateurs d'IC à améliorer leur perception auditive non verbale, comme précédemment suggéré avec du matériel verbal. Sur la base de cette étude, nous avons conçu une nouvelle stratégie d'entrainement à domicile utilisant l'intégration multisensorielle et plus spécifiquement la stimulation audiovisuelle. Nous suggérons que cet entrainement pourrait améliorer les capacités auditives non verbales des utilisateurs d'IC mais aussi des participants contrôles. Nous visons à démontrer l'efficacité de cet entrainement à long terme à la fois en utilisant des mesures comportementales (avec la batterie d'évaluation développée) chez toutes les populations avec un déficit,

mais aussi des mesures en magnétoencéphalographie chez les participants contrôles. Dans l'ensemble, ce travail de doctorat apporte des connaissances supplémentaires dans le domaine de la cognition auditive non verbale et des déficits associés, plus précisément dans le domaine de la perception des émotions. De plus, ce travail fournit un outil intéressant pour remédier à ces déficits.

Mots-clés : cognition auditive non-verbale, prosodie émotionnelle, émotions musicales, hauteurs tonales, connaissance implicite, intégration multisensorielle, entrainement audiovisuel, électroencéphalographie, entrainement, magnétoencéphalographie

Contents

ABSTRACT 7

RÉSUMÉ 9

INTRODUCTION 14

THEORETICAL BACKGROUND 16

I. MECHANISMS OF AUDITORY PERCEPTION AND COGNITION 17

A. Introduction 17

B. Definition and study of perception 17

1. Definition 17
2. Acoustic properties of sounds 17
3. Methods for the study of perception 18

C. Relationships between the anatomy and functions of the auditory system 20

1. Auditory perception on a psychological point of view 20
2. Auditory perception from an anatomo-functional point of view 22
3. Conclusion 27

D. Non-verbal auditory cognition 27

1. Definition of the non-verbal auditory cognition 27
2. Perception of emotions in music and language 28

E. Conclusion 31

II. NON-VERBAL AUDITORY PERCEPTION DEFICITS 33

A. Congenital amusia 33

1. Deficit in individuals with congenital amusia 33
2. Neural correlates of congenital amusia 34
3. Deficits of emotional perception in music and language in congenital amusia 34

B. Brain lesion and non-verbal auditory acquired deficits 35

1. Different origins of brain lesions 35
2. Acquired non-verbal auditory deficits following brain lesion 36

C. Cochlear implant users 38

1. Implant description 38

2. Limitations of the implant 42

3. Consequences for non-verbal auditory perception 42

D. Conclusion *43*

III. TRAINING AND BOOSTING NON-VERBAL AUDITORY PROCESSING IN HEALTHY AND PATHOLOGICAL POPULATIONS 44

A. Introduction *44*

B. Adult plasticity of the auditory system *44*

1. Auditory plasticity occurring during the adult life 44

2. Crossmodal plasticity 46

C. Training strategies in populations with non-verbal auditory deficit: the example of CI users *48*

1. Introduction 48

2. Musical training 48

3. Training using multisensorial integration 49

OBJECTIVES 52

EMPIRICAL PART 55

PART 1: EMOTION PERCEPTION IN NON-VERBAL AUDITORY DEFICITS 55

IV. EMOTIONAL PROSODY PERCEPTION IN CONGENITAL AMUSIA 56

A. Objectives *56*

B. Study 1: Emotional prosody in congenital amusia: Impaired and spared processes *56*

C. Study 2: Pre-attentive processing of neutral and emotional sounds in congenital amusia *74*

V. MUSICAL EMOTIONS PERCEPTION FOLLOWING A BRAIN LESION: IMPACT OF LESION SIDE 124

A. Objectives *124*

B. Study 3: Recognition of musical emotions and their perceived intensity after unilateral brain damage *124*

PART 2: BOOSTING NON-VERBAL AUDITORY PROCESSING: A NEW TRAINING STRATEGY 145

VI. REHABILITATION IN PATIENTS WITH COCHLEAR IMPLANTS. 146

A. Objectives *146*

B. Study 4: Rapid assessment of non-verbal auditory perception in normal-hearing participants and cochlear implant users *147*

C. Study 5: Building a training program to enhance non-verbal auditory abilities (in progress) *219*

GENERAL DISCUSSION **231**

VII. DISCUSSION AND PERSPECTIVES 232

A. Emotion perception in non-verbal auditory deficits *233*

1. Impaired perception of emotions in music and speech 233

2. Brain mechanisms of emotion perception 236

B. Training strategy for non-verbal auditory cognition using multisensory integration *238*

1. Assessing non-verbal auditory cognition deficits 238

2. Rehabilitation of non-verbal auditory perception and cognition 239

VIII. CONCLUSION 242

REFERENCES **243**

Introduction

In everyday life, we constantly face multiple sensory stimuli and we always have to adapt our behaviour to respond to these stimuli. In particular, the complex auditory environment is always challenging our ears and our brain to process every sound. However, it appears that normal-hearing listeners are evolving in this auditory environment quite easily and without too many difficulties despite the challenges that the brain is facing: hearing the correct target sound, ignoring the noise, correctly processing these sounds, and responding to it rapidly. Many different sounds can arrive to the ears, it can be environmental sounds (from natural sounds such as animal screams to human-created sounds such as horns), but also speech, and music. All of their spectro-temporal features are processed by the auditory system. However, auditory perception relies not only on the peripheral level (ear) but also requires higher order cognitive processing, such as learning, attention, and memory, that happens in cortical areas.

Non-verbal auditory cognition is essential in one individual's life to interact with the environment, including other people. For music, non-verbal auditory processing allows the listener to listen and make sense of series of sounds, to determine a melody and to recognize familiar pieces. For speech, non-verbal auditory cognition allows to determine the intentions and emotions of the speaker. In noisy situations without sufficient informative verbal cues, prior knowledge gives cues about auditory source characteristics and helps the listener to recognize this source.

However, non-verbal auditory perception can be degraded in some deficits, congenital or acquired, peripheral or central. For instance, congenital amusia is a central deficit of music perception and production, specifically related to pitch processing. Congenital amusic individuals have greater difficulties to recognize a familiar melody without speech and have shown deficits with pitch memory tasks. This deficit is present at birth and is not due to a brain damage. On the contrary, acquired disorder such as brain lesions can also affect non-verbal auditory perception. For instance, after a stroke, some patients report deficits for speech and music perception. These patients, as well as congenital amusics,

do not have any ear impairment that could explain their perception deficit. Auditory perception can also be a challenge for people with peripheral hearing loss. In the case of total deafness, cochlear implants are used to restore the hearing. However, cochlear implants users still experience great difficulties to perceive sounds, as only a limited number of frequency channels can be implanted and used in implants. Finally, in all of these disorders, auditory perception is degraded, and it causes several issues such as difficulties to hear in noisy context, but also deficits for understanding emotions of one speaker, and mostly a very limited music perception and enjoyment.

In this context, the present PhD thesis focuses on two main points. First, we aimed to better characterize emotion perception in music and speech in several non-verbal auditory perception deficits. We studied amusia (congenital and acquired) as a model of pitch perception deficit to study emotional perception. Second, we developed a new rehabilitation program to enhance this perception. To do so, we choose to focus our program on new cochlear implants users as they have high demands on retrieving the joy of listening to music.

Theoretical background

I. Mechanisms of auditory perception and cognition

A. Introduction

Here, we present the current state of the art of the human auditory system. First, we will review the study of perception for both its psychological and neurobiological functioning. Then, we will describe the auditory system.

B. Definition and study of perception

1. Definition

Perception allows us to acquire knowledge about the environment. Information is received by sensory systems (visual, touch, auditory, taste, smell) and then processed by the perceptive systems to make sense of this sensory information. Usually, perception is considered as an interpretative mechanism, allowing perceivers to build a mental representation of the environment from incomplete information. Auditory information processing can depend on both bottom-up processing (the features in the acoustic environment) and top-down processing (prior knowledge, attention, but also the internal state of the auditory system of the individual).

2. Acoustic properties of sounds

Sounds are vibrations of the air that propagate up to the ears (Grondin, 2016). These vibrations are a succession of compressions and rarefactions of the air (rising and falling of the pressure). The physical properties of a sound depend on these changes of pressure across time. The speed of these changes (number of cycles per second) is called the frequency of the sound, measured in Hertz (Hz). The pressure variations can be more or less pronounced which determines the amplitude (or intensity) of the sound, measured in decibel (dB). Complex sounds are composed of multiple waves of different frequencies. In these sounds, the lowest frequency is called fundamental frequency (F0) or first harmonic. A sound can be considered harmonic when it only contains harmonics of the F0, otherwise it is considered inharmonic. A harmonic corresponds to a wave with a frequency which is an integer multiple of the F0. The relative importance of each harmonic in a sound can also change the perceptual quality, for instance,

one note played by two different instruments. For example, one note played by a piano and a violin will have the same F0 but will be perceived as different by the listener.

The physical properties of the sounds are interpreted by the human ear and brain, and each sound can evoke different impressions in the listener. These define the subjective characteristics of the sounds. One of them is *pitch,* which is closely related to the frequency of a sound. Pitch is the perception of a sound that allows to order it on a frequency-related scale. Specifically, it is the judgement of a sound as high or low (Plack et al., 2005). It depends on the low or high frequencies composing the sound, but also on the intensity of the sound. Pitch tends to be perceived lower when the intensity is increased (Snow, 1936). Loudness is also a subjective dimension of auditory perception and is directly linked to the intensity of a sound but also to its frequency. Loudness tends to be perceived lower when frequency increases. Finally, a subjective dimension of the perception of a sound is called the timbre. This dimension depends on the composition of each sound in terms of frequencies and harmonics and is involved in the perception of different music instruments.

3. Methods for the study of perception

Perception can be studied according to two main angles. On one hand, psychophysical and cognitive psychology studies perception with behavioural measurements. On the other hand, neuroscience investigates the cerebral correlates of perception. These brain correlates are mostly studied with electrophysiology and imagery in the human brain.

a. Psychophysics and cognitive psychology

Experimental psychology studies mental abilities, and among them, perception. In this approach, psychophysics allows for quantitative analyses of the relationship between physical characteristics of the stimulus and the evoked sensation; whereas cognitive psychology decomposes different mental abilities that allow a behaviour in fine.

Psychoacoustics, issued from psychophysics, studies the relationship between the physical characteristics of the sound and the evoked sensation after hearing a sound. This discipline studies various phenomena, such as perception threshold against silence for stimuli but also hearing in noise

and masking situations. Generally, this research area allows for the understanding of the perceptual system's pre-processing of auditory stimuli in the nervous system (Moore, 2003).

Auditory perception goes far more beyond the basic mechanisms described by psychoacoustics (McAdams & Bigand, 1993). To better understand auditory perception, in particular sound recognition or music perception, several cognitive factors are needed, such as memory or attention. Auditory cognition studies this auditory perception in the framework of the personal experience of the listener, taking in account his/her previous knowledge but also his/her desire to better understand its environment. In cognitive psychology, usually the process of perception is represented with several levels (or steps) of processing. Each of these levels has a specific function and these levels can have several relationships between each other.

b. Brain activity measures

Brain activity corresponds to the information transfer between neurons. This information is transmitted in the shape of action potentials from the synapse of one neuron to the dendrites of one (or several) neuron(s). This information is then transmitted to the cellular body of the neuron. To process a sensory information or to execute a task, the brain needs the coordinated activity of neurons of one (or several) specific region(s). Hence, various measures allow to understand this brain activity at several levels:

- Electrical measures at the cellular level: measuring the pre and post-synaptic membrane potentials by using animal electrophysiology.
- Electrical measures at the brain level: measuring the consequences of the combination of cellular electric activities thanks to human electrophysiology. These tools can measure electric current (electroencephalography (EEG)) or magnetic fields (magnetoencephalography (MEG)) at the scalp level.
- Metabolic measures of the functioning of the brain: measuring the neurotransmitters changes (production, delivering) at the synapse level, or the variation of blood oxygen level at the brain region level (functional magnetic resonance imagery (fMRI)).

- Neuropsychology with brain-injured patients: coupling with magnetic resonance imagery (MRI) of the patient's brain to late the lesion, the correlation with their behaviour allow to better understand brain region's functioning.

Several methods have been developed over decades to better understand the brain functioning, with more or less precision based on the invasiveness degree of the methodology. These methodologies combined with the study of perception have allowed more precise determination of how auditory signals are processed over the auditory pathway from the ear through the brain.

C. Relationships between the anatomy and functions of the auditory system

Two main ways of studying auditory perception can be distinguished: (1) psychoacoustics and psychology study the different levels of sound processing (Grondin, 2016) from the encoding to the understanding (language, music, auditory source), with the extraction of the sound characteristics. (2) neurobiology studies the regions involved in the process of the sound with their connections (Ehret & Romand, 1997; Grondin, 2016), from the cochlea, through the auditory nerve, to the subcortical and cortical brain networks (see **figure 1**).

Clearly, these two ways of studying are complementary and can correlate with one another on several levels. However, it is quite difficult to associate one level of sound processing, as observed in cognitive psychology studies, with one region of the auditory system, as several levels of sound processing can be done within the same region, and several regions are needed for the processing of one type of sound. Hence, here we will explain these two ways of studying the auditory system by describing their processing levels. Moreover, we will describe the ascending and descending connections between these processing levels.

1. Auditory perception on a psychological point of view

Auditory perception relies on the separation of the different auditory sources coming to the ears and on the recognition of these sources. All of these processes are called auditory scene analysis (ASA).

In this part, we will focus on understanding how sounds are identified as speech, music or even just noise, but also how the auditory system allows to distinguish several sounds coming at one time to the ears (Grondin, 2016; McAdams, 2015; McAdams & Bigand, 1993).

a. Streaming

As Bregman (1994) demonstrated, the first step of the analysis of an auditory scene (ASA) is either the segregation of the sensory information into several flux, or the integration of several elements into the same flux, the same "sound object". A stream then refers to the perceptual unit forming an object. Bregman suggested that two principal mechanisms could allow the ASA: a primitive analysis or an analysis based on schemes.

The primitive analysis is based on acoustic properties of sounds and allows distinguishing several auditory sources without prior knowledge about them. For instance, close pitches of different sounds, but also their rhythmic pattern over time, can give cues to form streams. The classic example is the Van Noorden effect (Van Noorden, 1975): if two sources have two frequencies far from each other, they will be perceived as two segregated streams, but if these two frequencies are close to each other, they become part of the same stream. Moreover, this perception is modulated by the speed presentation of sounds, the slower the tempo is, the more segregate the two streams will be.

Compared to the primitive analysis, the analysis based on schemata suggest that stream segregation can be partially constrained by the knowledge and the expectations of the listener. According to the context, we could expect more specifically one type of sounds (our name for example), and would use existing schemes to analyse the incoming auditory sources.

All of these mechanisms are essential to distinguish or group the sensory information along time. They allow determining several streams that can then be categorized and understood by the listener according to their type of sounds (speech, environmental sound, music, etc.).

b. Recognizing the auditory source

Recognizing the type of auditory source relies on acoustic cues as well as previous knowledge of the listener. However, determining which acoustic cues can be attributed to one type of auditory source is

not trivial as a same auditory object can emit different sounds according to the context. Acoustic invariants associated to each type of sound mainly relate to frequency, intensity and temporal but no real consensus has yet been reached on the specific nature of these invariants for each sound (Grondin, 2016; Handel, 1989).

In music, chaining different notes varying in pitch forms the melody. A melodic phrase is perceived as one harmonious stream and not as a succession of independent, individual notes. In Western tonal music, these variations of pitches rely on a specific structure, a chromatic scale. A note is associated to a specific pitch in an octave, which is composed of 12 semitones. Ratios between pitches are constrained by these octaves. Rhythmicity of a musical excerpt is also determined by the specific duration of the notes (quarter, half, whole). All of these coherent variations are perceived as a whole and contribute to the percept of a melody. Altogether, it allows the enjoyment of music, sometimes influenced by the personal experience of the listener as well as her/his memory of melodies.

Speech sounds are distinguished according to their acoustic cues, described by the field of phonetics. Units of language, called phonemes, are put together to make meaningful sounds. Each language had its specific phonemes and different rules to pronounce them. Here again, perception and understanding of speech rely mainly on the knowledge of the listener, especially to reconstruct sentences from isolated words and to make sense of these words. In speech, sounds can vary according to the accent, the context or even the age of the speaker. However, despite these variations, the listener is usually still able to make sense of these sounds.

2. *Auditory perception from an anatomo-functional point of view*

From the neurobiology point of view, the auditory processing starts when the sound arrives to the ear, in the cochlea and it is transmitted up to the auditory cortex (AC) and beyond. This ascending pathway is also under the control of numerous descending pathways that will not be described in detail in this presentation. Here, we will give insights to understand the role of each step along the auditory pathway in each anatomical region. These results come mostly from data in the animal domain, anatomical data and lesions data in humans (MRI). For details, see Grondin (2016), Ehret&Romand (1997), **figure 1**.

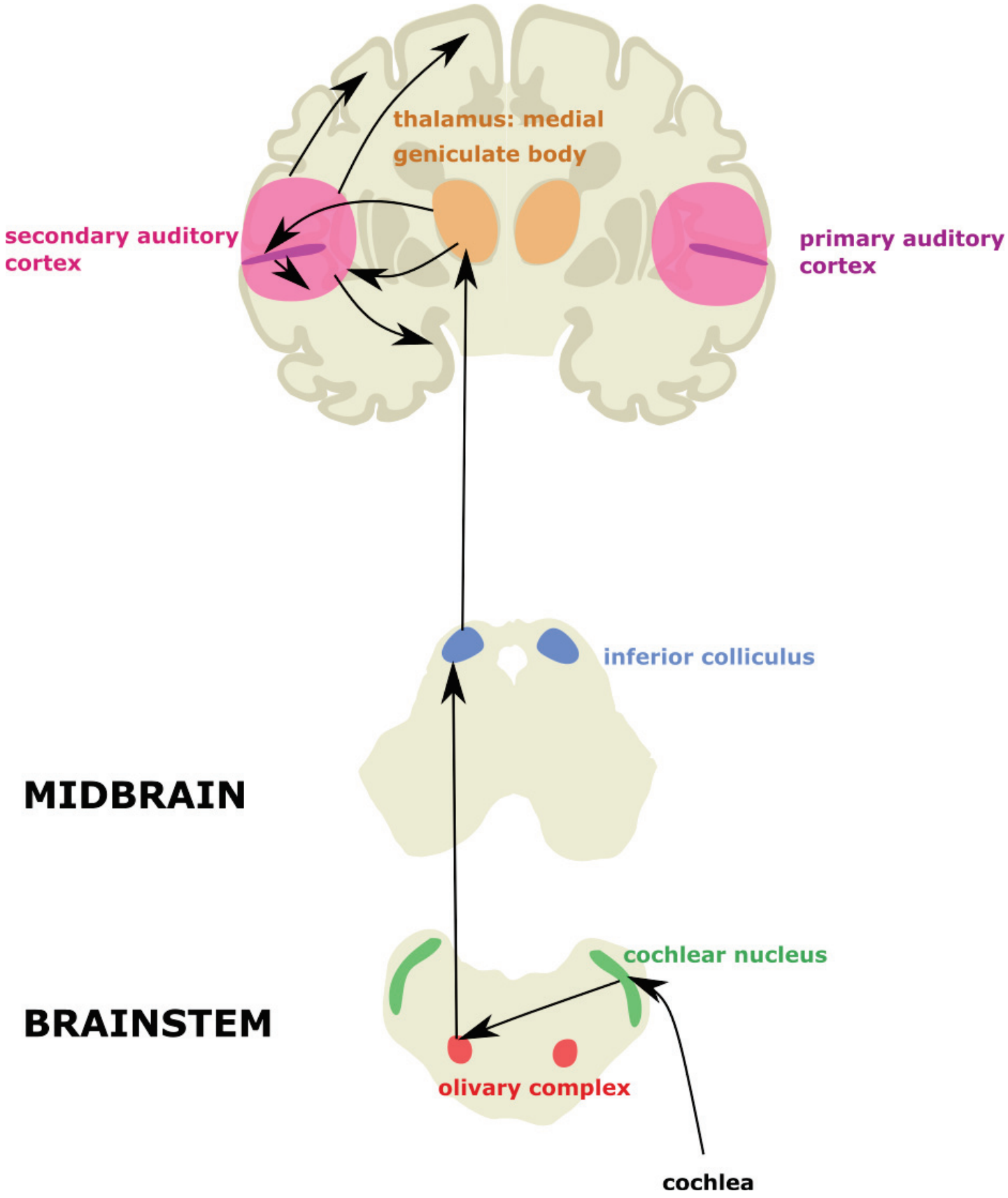

Figure 1: *Ascending auditory pathway, from the cochlea to the cerebral cortex. Main projections between the regions are represented by arrows.*

a. In the cochlea

The various acoustic messages for several auditory sources pass through the outer and middle ear and arrive at the inner ear when they reach the cochlea, where the auditory signal is filtered according to frequencies. At the centre of the cochlea, on the basilar membrane, at the organ of Corti, thousands of hair cells transform the acoustic signal into a neuronal signal transmitted to the auditory nerve fibres. The location of the hair cell on the basilar membrane determines its response to one frequency. Each region of the cochlea is dedicated to the analysis of one frequency, as if the cochlea is applying pass-band filters on the acoustic signal. High frequencies are analysed by the hair cells of the base of the cochlea whereas low frequencies are analysed by the apex of the cochlea (see **figure 2**).

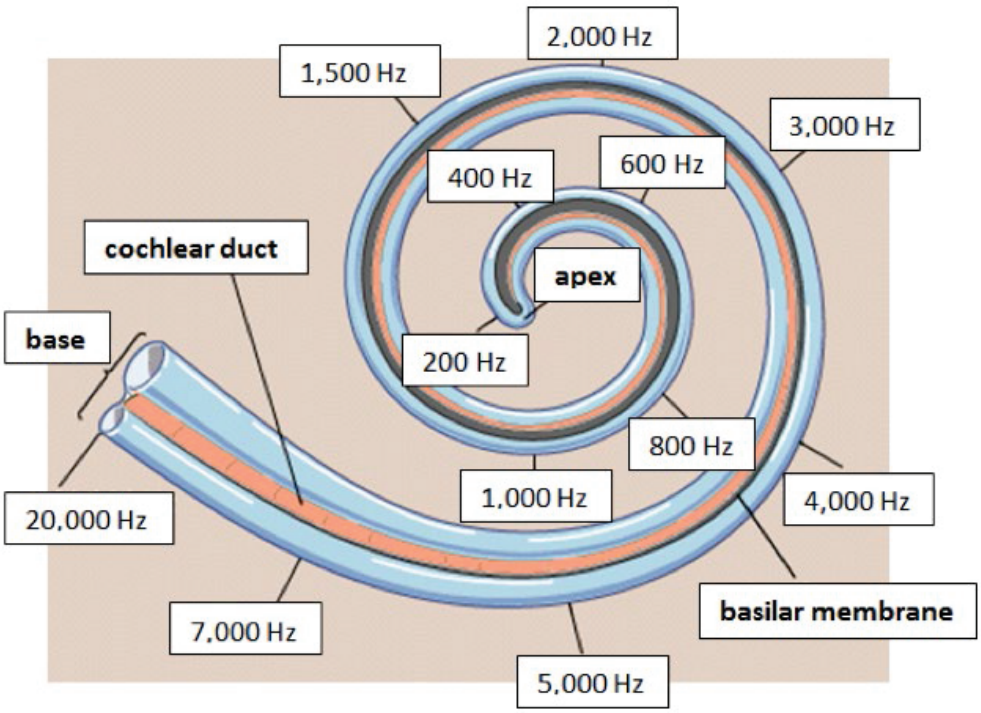

Figure 2: *cochlear organization and hair cells' responses to specific frequencies. From Encyclopedia Britannica, http://www.britannica.com/EBchecked/media/537*

The overall signal sent to the auditory nerve is the combination of the frequency analysis made by all hair cells. In the auditory nerve, each fibre also has its own specific frequency. This tonotopy has been observed in the auditory subcortical structures and up to primary auditory areas in the cortex.

b. Subcortical auditory structures

Information from both ears arrives in these structures and is processed. Auditory nerve fibres arrive at the ventral and dorsal cochlear nucleus. Neurons of the ventral nucleus project to the superior olivary complex. In this complex, spatial information about the stimulus is processed, based on the intensity differences and the arrival time differences between the two ears. Neurons of the dorsal nucleus and the olivary complex project to the inferior colliculus. Three pathways are going from the inferior colliculus to the cortex, passing by the thalamus (including the medial geniculate body): a tonotopic pathway from central inferior colliculus to the thalamus and the primary AC; two pathways from the colliculus to the secondary AC via the thalamus. One pathway is auditory, the other is multisensory, combining somato-sensory information. Each neuron of the subcortical auditory structures is sensitive to more and more complex features of the stimulus, such as frequency modulations or intensity modulations.

c. The auditory cortex

The auditory cortex (AC) is located in the Sylvian fissure in the supratemporal plane, and goes from the superior temporal gyrus to the parietal cortex (Galaburda & Sanides, 1980; Rivier & Clarke, 1997). The primary AC is in the middle of the Heschl gyrus, surrounded by secondary auditory regions called the parabelt. The primary AC is organized tonotopically with frequency lines along the Heschl gyrus axis, from low frequencies at the external gyrus to high frequencies at the centre (C. Pantev et al., 1995). The central zone of the primary AC corresponds to the frequencies of formants of speech signal.

Associative auditory cortices surround the primary AC (Galaburda & Sanides, 1980; Rivier & Clarke, 1997; Wallace et al., 2002). Based on fMRI, EEG and MEG data, four tonotopic regions can be distinguished (C. Pantev et al., 1995; Talavage et al., 2000), two of them processing only low frequencies.

The processing of sound in these cortices is hierarchical from the cochlea to the primary areas and the secondary areas (Rauschecker et al., 1997). In the secondary areas, hierarchical processing is also observed within the belt and the parabelt (Kaas et al., 1999). However, the sensory information is also processed simultaneously in several regions of the AC, depending on the characteristics of the sounds.

d. Beyond the auditory cortex

Two main pathways were identified to process the auditory information at the cortical level beyond the AC: the "what" pathway (ventral), the "where" pathway (dorsal) (Bizley & Cohen, 2013), similarly as what is observed in the visual domain. The dorsal pathway goes from the posterior temporal auditory regions to the dorsolateral prefrontal cortex via the parietal regions. The pathway is involved in the processing of the spatial information of the stimulus and thus is involved in audio-motor processing, as well as speech perception and production (Rauschecker, 2018). Whereas the ventral pathway goes from the anterior regions of the AC and projects to the anterior superior temporal gyrus and the ventral prefrontal cortex. This pathway is involved in the auditory-object processing (Bizley & Cohen, 2013). Even if they give a good insight about the higher processing of auditory sources, other brains regions are also involved in those processes such as the hippocampus. These regions would be more involved in attention and memory processes of the auditory information but their roles on the analysis of the auditory scene are still to be investigated (Bizley & Cohen, 2013).

e. Focus on pitch perception and its cerebral correlates

As pitch is essential for auditory cognition, the cerebral correlates of its perception were studied in detail. The pitch of a sound derives from the maximal activation along the tonotopic map, starting in the cochlea and throughout the tonotopic pathway described above, to reach the primary auditory cortex. In Heschl's gyrus, several frequency gradients are present along the antero-posterior axis and the medio-lateral axis, and contribute to pitch perception (Formisano et al., 2003; Langers et al., 2007; Talavage et al., 2004). In secondary auditory cortices, perception of pitch seems to be tonotopically regionalized in the lateral and posterior superior temporal plane. Tonotopic gradients could also extend to the superior temporal gyrus (Saenz & Langers, 2014). Some studies on brain-damaged patients (Johnsrude et al., 2000; Stewart et al., 2006) and then fMRI studies, suggested that pitch perception could be mostly lateralized in the right auditory cortex. In particular, the right planum temporale activates in correlation with the size of frequency shifts between two tones (Hyde et al., 2008). In addition, as music and speech are usually sequences of multiple pitches, some studies have looked at the effect of varying pitch in sequences on the brain activity. They found that these sequences activated areas in the superior temporal lobe but also

more distant brain regions such as the inferior frontal areas (Griffiths, 1999; Griffiths & Green, 1999; Janata, Birk, et al., 2002; Janata, Tillmann, et al., 2002; Koelsch et al., 2009; Peretz & Zatorre, 2005; Schulze et al., 2011; Schulze et al., 2011; Tillmann et al., 2006; Tillmann et al., 2003; Zatorre et al., 1994; Zatorre et al., 2002). Moreover, specific regions were also demonstrated to be involved for short-term memory for pitch. In addition to auditory cortices, neuroimaging studies demonstrated activation of a specific cortical network for this memory (Foster et al., 2013; Foster & Zatorre, 2010; Gaab et al., 2003; Griffiths & Green, 1999; Zatorre et al., 1994). In particular, inferior frontal and insular cortex, supramarginal gyrus, the planum temporale and cerebellar regions would be involved in pitch retention and pitch maintenance processes. These studies suggested that short-term-memory for pitch would require interactions between frontal and posterior temporal brain networks.

3. Conclusion

From the ear to the auditory cortices and beyond, auditory sources are processed and identified. Several regions process the sound according to its frequency but also to its intensity or its variations across time. In the central nervous system, information from both ears are combined and the involvement of several other regions apart from the AC is important to understand the sounds and the auditory scene. These processes are essential for the perception of speech allowing the human being to communicate and exchange with others. However, other non-verbal auditory cues are essential to understand the auditory scene and make sense of the environment.

D. Non-verbal auditory cognition

1. Definition of the non-verbal auditory cognition

Non-verbal auditory cognition comprises all auditory processing of sound that does not involve directly the understanding of language. Non-verbal auditory processing allows for the perception of emotions and the intentions of the speaker in speech (called emotional and intentional prosody), the localization of non-verbal sounds but also the perception of music. More generally, this cognition comprises all the cues to analyze the auditory scene without focusing on words. Perception of emotional content of a sound has specific networks in the brain. We will further describe these networks here.

2. *Perception of emotions in music and language*

a. *What is an emotion?*

The first point of complexity about emotions is the variety of their categories. Different ways of classifying what kind of emotions can be encountered when listening to sounds have emerged. Numerous studies have focused on basic emotions (a limited number of emotions) such as happiness, sadness, fear and peacefulness (Fritz et al., 2009). Emotions can also involve other affective states, such as nostalgia and power, to enhance the precision of expression of feelings (Vuilleumier & Trost, 2015). The classification of emotions according to categories are used in experiments, and offers a restricted number of choices for the participants (Omigie, 2016). Another possibility to categorize emotion is to use a two-dimensional representation, with axes corresponding to valence and arousal: Valence relates to the value of the emotion felt, it can be positive (attractive) or negative (repulsive); Arousal relates to the intensity of the emotions felt during listening (Bestelmeyer et al., 2017).

Many features of sounds can evoke emotions (Koelsch, 2015). The components of the sound itself can evoke various emotions. For example, a melody with a fast tempo is usually felt as joyful. The tension of music can lead to relaxation or a feeling of reward associated with fulfilling of what is expected by the listener (Spangmose et al., 2019). A direct link was found between the tempo of music and the entrainment of the listener, leading to change of arousal (Vuilleumier & Trost, 2015). Moreover, the knowledge and the emotional context of the listener can also influence the emotional value of a sound. For instance, the evaluation of the sound determines emotions felt in relation with the purpose of it. For example, music can be considered pleasurable if it is listened in order to feel energy. A sound can also be experienced as enjoyable or not in relation with the memory of the listener (Thompson et al., 2000). In particular, autobiographic memory and conditioning can lead to strong feelings when hearing a music (Koelsch, 2015; Omigie, 2016; Vuilleumier & Trost, 2015). However, it was also shown that sensitivity to music can be universal and does not always depend on the listener's experience. For instance, experiments of music emotion recognition in various cultural populations revealed that listeners can be sensitive to emotions even with unfamiliar tonal system (Balkwill & Thompson, 1999; Egermann et al., 2014; Fritz et al., 2009).

b. Shared networks for the perception of emotions

Emotional sounds are very complex and variating sounds. They can vary in term of acoustic content, but also in term of meaning. However, vocal and musical emotion appear to have share acoustic codes to detect them (Coutinho & Dibben, 2013; Nordström & Laukka, 2019). This is why, emotional prosody and music emotions still have common networks for their processing in the human brain (Escoffier et al., 2013; Frühholz et al., 2016; Sun et al., 2018). Both cortical and sub-cortical regions are involved in the decoding of emotions in sounds (see **figure 3**). The core neural network of emotional perception from sound signals comprises the limbic and auditory structures as well as frontal and subcortical regions (Frühholz et al., 2016; Koelsch, 2015).

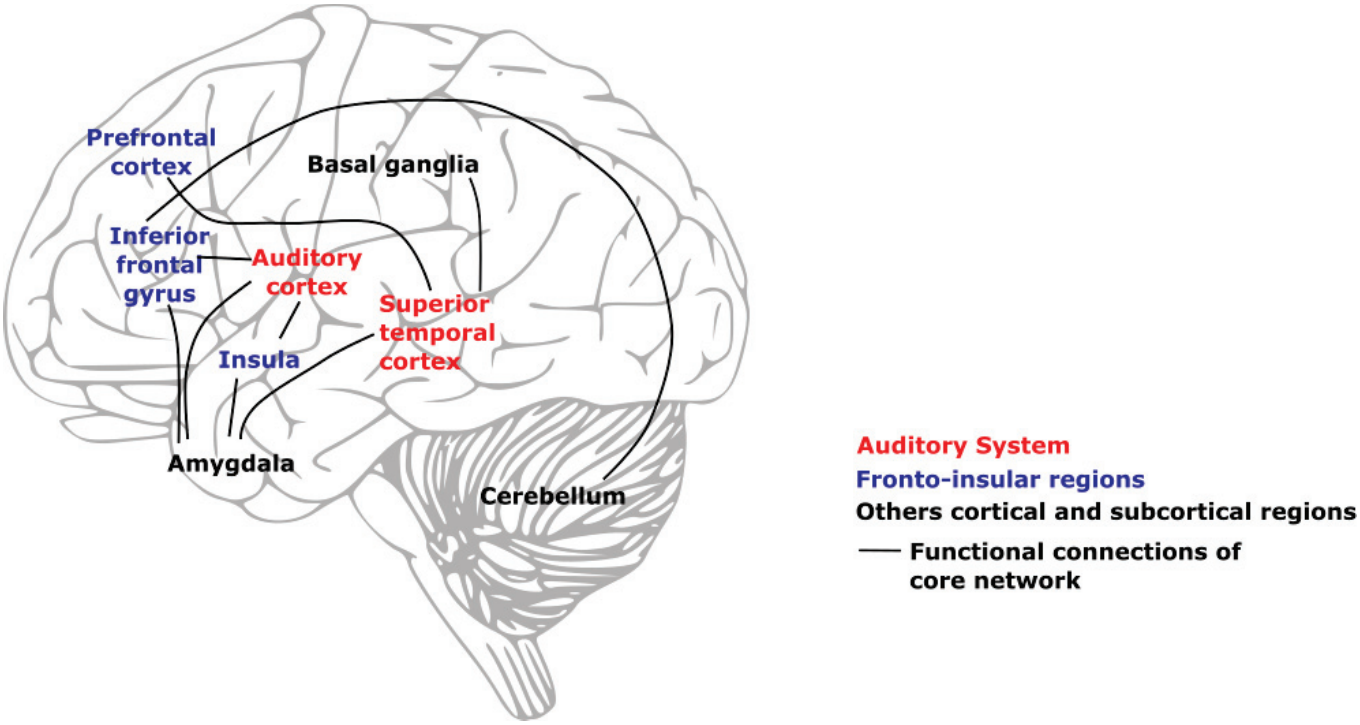

Figure 3: *Core brain networks involved in the processing of emotions. A core network is involved in every emotion listening, recognition and feeling. Red: ascending auditory pathway. Blue: fronto-insular regions. Black: other regions involved in emotional processing. Main functional connections between regions are indicated. Inspired from Frühholz (2016).*

The limbic system is well known to be essential for the processing of emotions, especially regarding the valence of emotion (Frühholz et al., 2015; Koelsch, 2014; Milesi et al., 2014). The amygdala is one of the key regions for the perception of emotional content in music and speech (Omigie, Dellacherie, Hasboun, Clément, et al., 2015; Omigie, Dellacherie, Hasboun, George, et al., 2015). More specifically, the amygdala receives signal from the medial geniculate nucleus and projects back to subcortical and

cortical auditory regions (Ball et al., 2007; Frühholz & Grandjean, 2013). The primary and secondary auditory cortices as well as the superior temporal cortex also play an important role to decode the emotional sounds (Frühholz & Grandjean, 2013; Koelsch, 2014; Kumar et al., 2012; Szameitat et al., 2010). Together, the amygdala and the auditory cortices process the emotional content of a sound and integrate this information into an auditory percept that can be shared with frontal regions.

Several frontal regions respond to emotional sounds. In particular, the inferior frontal cortex in association with the superior temporal cortex (via dorsal and ventral connections) process the relevant emotional sound and support its categorization according to its social meaning (Frühholz & Grandjean, 2012; Hoekert et al., 2008, 2010; Rauschecker & Scott, 2009). The insula, close to the inferior frontal cortex, also responds to emotional sounds (Frühholz et al., 2012; Kumar et al., 2012; Sander & Scheich, 2005; Trost et al., 2012). Specifically, the insula is involved in the detection of salience of emotional sounds, but also in the perception of this sound according to self-experience of emotions (Kotz et al., 2013; Wildgruber et al., 2004). The medial frontal cortex supports the processing of the social aspect of emotion, with its understanding in the context of communication (in link with memories and associations) (Amodio & Frith, 2006).

Basal ganglia and cerebellum also play a major role in the adaptation in response to the emotional perception of a sound (Sammler et al., 2015). Specifically, the ventral basal ganglia is involved in the processing of musical emotions and affective voices (Paulmann et al., 2005; Paulmann et al., 2008; Pell & Leonard, 2003). The dorsal basal ganglia also participates to the perception of emotions as it contributes to their decoding over time (Kotz & Schwartze, 2010; Salimpoor et al., 2011). Indeed, the dorsal ganglia helps to decode temporal patterns in emotional prosody as well as temporal anticipations in music (Grahn & Brett, 2007; Hass & Herrmann, 2012). This decoding is done in association with the cerebellum (Ethofer et al., 2012; Kotz et al., 2013). In particular, this region is sensitive to sudden emotional changes and can lead motor responses to aversive sounds (Zald & Pardo, 2002) but also to the motor reaction to music (Trost et al., 2014).

In conclusion, a large core network is involved in the perception and processing of emotional sounds, whatever their types. In addition to this core network, other specific brain regions can be involved in this processing such as the hippocampus for the association with the individual's episodic memory. Moreover, each affective sound (especially music or prosody) is processed preferentially by specific regions that we will describe in detail.

c. *Brain networks for music emotion perception*

Musical emotion processing models were mainly elaborated based on anatomical connectivity in the limbic system, in particular the temporal and frontal lobe and the limbic striatum, as well as functional MRI data (Koelsch, 2014). Specifically, in the medial frontal cortex, musical emotions seems to activate more strongly the ventro-rostral part (Frühholz et al., 2016). Ventral striatum and basal ganglia respond to musical emotions, especially in the nucleus accumbens. This suggests that listening to music can be associated to reward and pleasure (Belfi & Loui, 2019; Koelsch, 2015). Moreover, the hippocampus and the orbitofrontal cortex have strong connections that respond to complex emotions, especially in relation to listeners' long-term memory.

d. *Brain networks for emotional prosody perception*

Models of emotional voice processing generally comprise two hierarchical pathways. One feedforward pathway comprises processing in the auditory-inferior frontal cortex; the other pathway comprises processing in the amygdala and in the auditory-inferior frontal cortex (Schirmer & Kotz, 2006; Wildgruber et al., 2009). The cerebral correlates of prosody processing involve bilateral inferior frontal gyri (Frühholz et al., 2012). Specifically, in the medial frontal cortex, emotional prosody seems to activate more strongly the dorso-caudal part (Frühholz et al., 2016). In addition, emotional prosody processing involves the right anterior superior temporal gyrus (Frühholz et al., 2012; Liebenthal et al., 2016; Sammler et al., 2015).

E. Conclusion

Auditory perception is essential to understand the environment and react to it. Speech communication is specific to humans and allows for efficient communication between individuals. However, non-verbal

auditory cognition is also determinant for this communication and sensory processing. It allows for the comprehension of auditory sources and the analysis of auditory scene. These processes are present from the very beginning of life, as babies are able to process and appreciate emotions in voices but also in music. In some deficits, this non-verbal processing is disrupted and can degrade the perception the auditory environment of people suffering from this deficit.

II. Non-verbal auditory perception deficits

The mechanisms of auditory perception and the specific perception of non-verbal aspects of sounds has been reported as impaired or less efficient as in healthy individuals and lead to deficits of communication or of perception of the auditory environment. Here, we will focus on three specific deficits of non-verbal auditory processing. This first deficit is a central deficit of music perception, specifically related to pitch processing and referred to as congenital amusia (CA). This deficit is present at birth and is not due to a brain damage (part 1). On the contrary, acquired disorder such as brain lesions can also affect non-verbal auditory perception (part 2). However, no ear damage is present in any of these two deficits. Finally, non-verbal auditory perception can also be a challenge for people with hearing loss and wearing cochlear implant(s) for hearing restoration. As hearing loss can be present from birth or occur later in life, here we will focus on previous work done with adult cochlear implants' users (part 3).

A. Congenital amusia

1. Deficit in individuals with congenital amusia

CA is a lifelong deficit of music perception and production, also referred to as "tone deafness". This disorder is estimated to affect one to four percent of the general population (Peretz et al., 2007; Peretz & Vuvan, 2017) and is suggested to have genetic origins (Peretz et al., 2007). Amusic individuals have neither peripheral auditory deficits nor brain lesions, but they are unable to detect out-of-key notes in a melody, and sing out-of-tune (Peretz, 2016; Barbara Tillmann et al., 2015). Pitch processing deficits were observed for amusic individuals in perception tasks, such as pitch discrimination (Hyde & Peretz, 2004) or pitch contour change detection (Peretz et al., 2003), as well as in short-term memory related to pitch (Graves et al., 2019; Tillmann, Lévêque, et al., 2016). The deficit extends to timbre (Stewart, 2011; Barbara Tillmann et al., 2009), whereas temporal processing seems to be mostly preserved in amusia (Hyde & Peretz, 2004), at least when the material does not entail pitch variations (Foxton et al., 2006; Pfeuty & Peretz, 2010).

As pitch processing is involved not only in music processing, but also in speech processing, several studies have focused on speech perception abilities in amusia. Interestingly, while early studies did not report deficits in speech processing (Ayotte et al., 2002) or in memory for verbal sounds (Barbara Tillmann et al., 2009; Williamson & Stewart, 2010), subsequent studies using more fine-grained materials and methods did reveal speech processing impairments in amusia. Specifically, intonation recognition and perception of speech contour is impaired across languages - this includes tonal languages, non-tonal languages, and even artificial verbal materials (Jiang et al., 2010, 2012; Liu et al., 2015, 2017; Nan et al., 2016; Nguyen et al., 2009; Barbara Tillmann, Burnham, et al., 2011; Barbara Tillmann, Rusconi, et al., 2011).

2. *Neural correlates of congenital amusia*

In the typical (non-amusic) brain, a fronto-temporal network is involved in pitch processing and memory, and thus also in music perception (Gaab et al., 2003; Koelsch et al., 2009; R. J. Zatorre et al., 1994). In the amusic brain, anatomical and functional abnormalities have been observed in this fronto-temporal network (Albouy, Mattout, et al., 2013; Hyde et al., 2006, 2007, 2011). More specifically, decreased fronto-temporal connectivity was observed in congenital amusia, in particular in the right hemisphere, together with an increased connectivity between the auditory cortices (Albouy, Mattout, et al., 2013, 2015; Hyde et al., 2011; Leveque et al., 2016; Loui et al., 2009; Barbara Tillmann et al., 2015). These findings suggest an altered auditory neural network underlying the pitch processing deficit in congenital amusia (Leveque et al., 2016).

3. *Deficits of emotional perception in music and language in congenital amusia*

Emotion processing has been studied with musical material in congenital amusia. Despite impaired perception and memory of music, some listeners afflicted with congenital amusia have been reported to either like or avoid listening to music (Mcdonald & Stewart, 2008; Omigie et al., 2012). This dichotomy occurs independently of the severity of amusia, as measured by the Montreal Battery for the Evaluation of Amusia (Mcdonald & Stewart, 2008; Omigie et al., 2012). These subjective reports about liking/avoidance inspired recent studies investigating musical emotion processing in congenital amusia.

Gosselin et al. (2015) showed no impairment of emotion recognition (Gosselin et al., 2015), while Leveque et al. found a mild impairment (Lévêque et al., 2018). Similarly, a study that focused on dissonance/consonance judgments of musical materials reported that congenital amusics were able to recognize the suggested musical emotions, but they based their judgments more on roughness rather than on the harmonicity cues used by control participants (Marin et al., 2015). These findings and previous reports of the perceptual deficits in amusia suggest that amusics' emotional judgments in music are based largely on roughness and tempo rather than harmonicity and mode cues (Gosselin et al., 2015; Lévêque et al., 2018).

As the right superior temporal and inferior frontal regions participate in emotional prosody processing, and these regions exhibit differences in amusia compared to controls (Albouy, Mattout, et al., 2013; Hyde et al., 2007), Liu et al. (2015) suggested that impaired processing in these regions could underlie amusics' difficulties processing subtle emotional and prosodic changes (Ayotte et al., 2002; Patel et al., 2005). Previous studies have demonstrated an impairment of emotional prosody recognition in amusics (Lima et al., 2016; Thompson et al., 2012,), especially for happiness, tenderness, irritation, and sadness (William Forde Thompson et al., 2012). However, this deficit was not found in every congenital amusics (Lolli et al., 2015).

Overall, CA is an interesting deficit to study the influence of impaired pitch perception on non-verbal auditory processing, without brain or ear lesions. The lifelong deficit and its study contribute to a better understanding of the complex relationship between music and emotional prosody processing, and provides further elements for the comprehension of fine acoustic structures underlying music and speech appraisal.

B. Brain lesion and non-verbal auditory acquired deficits

1. Different origins of brain lesions

Brain injuries are very diverse and can be either due to an accident, following a disease or after a surgery (Elbaum & Benson, 2007). In pharmaco-resistant epileptic patients, surgery can be proposed, and brain regions are removed to relieve the patients. Similarly, following a brain tumour, surgery can also be

proposed. Brain lesions can also appear following a stroke and an ischemia, generally in the elderly. These brain damages are various and often lead to multiple cognitive deficits. Brain-damaged (BD) patients have been very valuable to the understanding of brain organization and brain functioning. More specifically, brain-damaged patients' behaviour has been studied to better understand music and speech processing.

2. *Acquired non-verbal auditory deficits following brain lesion*

After a brain injury, specific cognitive impairments can occur affecting non-verbal auditory cognition. Studies of non-musician patients have explored the link between brain lesions and music processing (see Stewart et al., 2006 for a review). This has led to the identification of two mains impairments with music in these patients (Stewart et al., 2006): a deficit of music perception (called acquired amusia) (Clark et al., 2015; Hirel et al., 2014), and/or a deficit of musical emotion processing in term of emotion recognition or emotion feeling (called musical anhedonia) (Belfi et al., 2017; Hirel et al., 2014). These two deficits could occur conjointly or separately.

a. *Acquired amusia*

In the majority of deficits regarding music reported in BD patients, basic perceptual attributes (pitch, timbre, temporal cues) of music sounds cannot be processed correctly (Tillmann et al., 2017), leading to acquired amusia. In acquired amusia cases, pitch processing is altered (Sihvonen et al., 2016; Sihvonen, Ripollés, Rodríguez-Fornells, et al., 2017; Sihvonen, Ripollés, Särkämö, et al., 2017). More specifically, it can be a deficit of pitch interval or pitch contour patterns perception (Clark et al., 2015). Moreover, specific pitch short-term memory can also be deteriorated following a brain lesion, specifically in the right anterior temporal lobe (Tillmann et al., 2017 for a review). Regarding timbre, several studies have also demonstrated impaired processing of spectral or temporal dimensions of timbre in BD patients (Samson et al., 2002). Regarding rhythm perception in BD patients, some deficits can be detected as well (Clark et al., 2015).

Acquired amusia has been linked to damages in several brain regions. Specifically, amusia was reported in BD patients with lesions in the right superior temporal gyrus, the Heschl's gyrus, the middle temporal

gyrus, the insula and the putamen (Sihvonen, Ripollés, Särkämö, et al., 2017; Stewart et al., 2006 for reviews). Interestingly, acquired amusia can disappear in some cases in few months following the lesion, but persistent amusia remains in BD patients and was associated with a deficit of grey matter volume in the right superior temporal gyrus and the middle temporal gyrus, and a deficit of white matter volume in the middle temporal gyrus (Sihvonen, Ripollés, Särkämö, et al., 2017).

b. Emotional music perception deficit

Several case reports have revealed that brain damage can result in musical anhedonia, a specific loss of experience of pleasure for music whereas emotion recognition is intact (Belfi et al., 2017; T. D. Griffiths et al., 2004; Satoh et al., 2011, 2016). For instance, a patient with a right inferior parietal lobe infarct did not experience any emotion when listening to music, but music perception and emotion recognition were preserved (Satoh et al., 2011). Deficits of music emotion recognition in brain-damaged patients have been shown in case reports (Gosselin et al., 2007; Griffiths et al., 2004; Satoh et al., 2016) and group studies (Gosselin et al., 2011; Jafari et al., 2017; Khalfa et al., 2007, 2008) . For instance, patients with unilateral medial temporal lesions showed more difficulties to recognize musical emotions, especially fearful stimuli (Gosselin et al., 2011). However, when asked to judge emotional dissimilarities in musical excerpts in terms of arousal and valence instead of emotion recognition, patients with left or right unilateral medial temporal lesions did not show any deficit (Dellacherie et al., 2011). Interestingly, it has been shown that emotion recognition can be preserved in right brain-damaged (RBD) patients even when their musical structure perception is impaired (Peretz et al., 1998), suggesting that acquired amusia and acquired anhedonia can be dissociated. The variety of profiles observed among these patients are in keeping with the hypothesis of (at least partly) separate processes for music perception and emotion (Peretz et al., 1998; Satoh et al., 2011; Stewart et al., 2006). These results suggest a complex pattern of musical emotion perception in BD patients with remaining work to do to understand it.

c. Emotional prosody deficit

As aphasia can occur in many brain-lesioned patients, the processing of other speech related contents was studied in these patients. Specifically, recognition of emotional prosody was mainly studied in BD patients as it allows for communication even without good speech perception abilities (Yuvaraj et al.,

2013). Studies using language and vocalization materials demonstrated emotional prosody deficit in RBD patients (Bourgeois–Vionnet et al., 2020; Kucharska-Pietura et al., 2003). More specifically, these studies report greater impairment of RBD patients for emotional prosody recognition than left brain-damaged (LBD) patients and controls (Borod et al., 2002; Charbonneau et al., 2003; Harciarek et al., 2006; Kucharska-Pietura, Phillips, Gernand, & David, 2003).

C. Cochlear implant users

1. Implant description

Cochlear implants (CI) was the first viable treatment elaborated to remediate deafness. At the beginning of implant conception, only profoundly or totally deaf persons were implanted. CI was only considered as a supply to hear environmental sounds and help to understand speech with lip reading. The first report of CI is dated of 1957 (Djourno & Eyries, 1957). It was a single electrode simulator on the auditory nerve. Patients could only recognize environmental sounds and few words. The first multiple electrode array appeared in 1966 (Simmons, 1966). It was implanted on the right auditory nerve. Patient could detect loudness related to the stimulus intensity and pitch related to the stimulus pulse rate. Separate pitch was associated to separate activation of six electrodes implanted. With only that kind on CI, patients could identify some well-known melody. In 1977, Bilger (1977) reported some results of 12 users of single electrode CI. This article claimed that patients could discriminate changes in low frequencies (below 250 Hz) only and could perceive loudness changes related to stimulus intensity. Early reports on CI users' perception don't describe any or very few music perception and recognition (Bilger, 1977; Eddington et al., 1978; Fourcin et al., 1979).

Cochlear implant technology has evolved over the past few years, but the basic functioning remains the stimulation of auditory neurons directly by electric currents. There were many different designs created, here we will describe general and common features for all of them (see **figure 4** for a general organization of the implant).

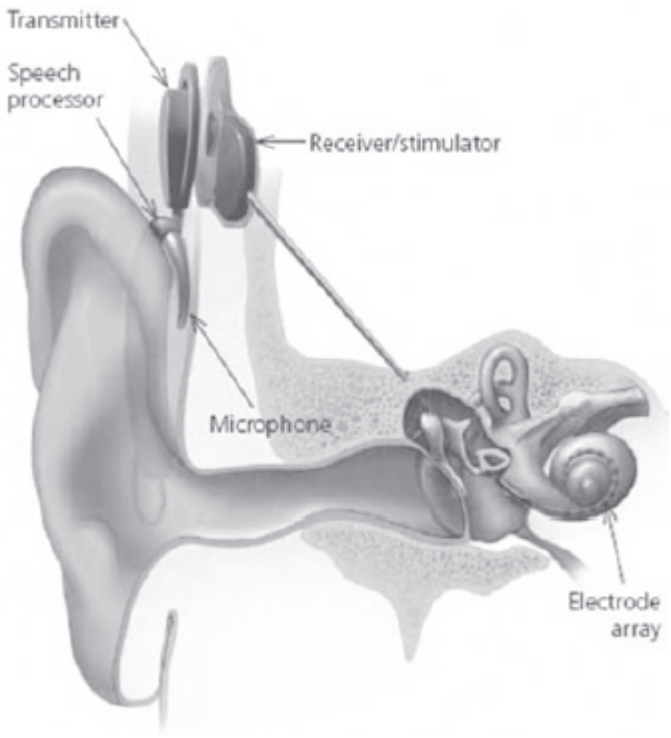

Figure 4: *Insertion of a cochlear implant in one ear, relationships between the different components. Figure from Hartmann (2013).*

The signal comes to a microphone, generally in the ear's pinna. This microphone transfers an electric signal to a signal processor. The electric signal corresponds to the transformation of soundwaves into variation of electric currents. The processor can convert these electric features to electric stimuli to the nerve that interprets these as proper hearing sensations. This process depends mostly on the algorithms used by the fabricant of processor (see Loizou (2006) for a review). Sound processor generates a code transmitted to a link; this link will provide electric power but also the correct electric stimuli to the implanted electrode array. The implant circuit then decodes this information from the sound processor. This step is crucial to determine the spectral and temporal parameters of the stimulus delivered by the implant; this decides which electrodes will conduct the stimulus current. Implants can generate biphasic pulses or continuous varying currents (analog). See **figure 5** for a functioning scheme.

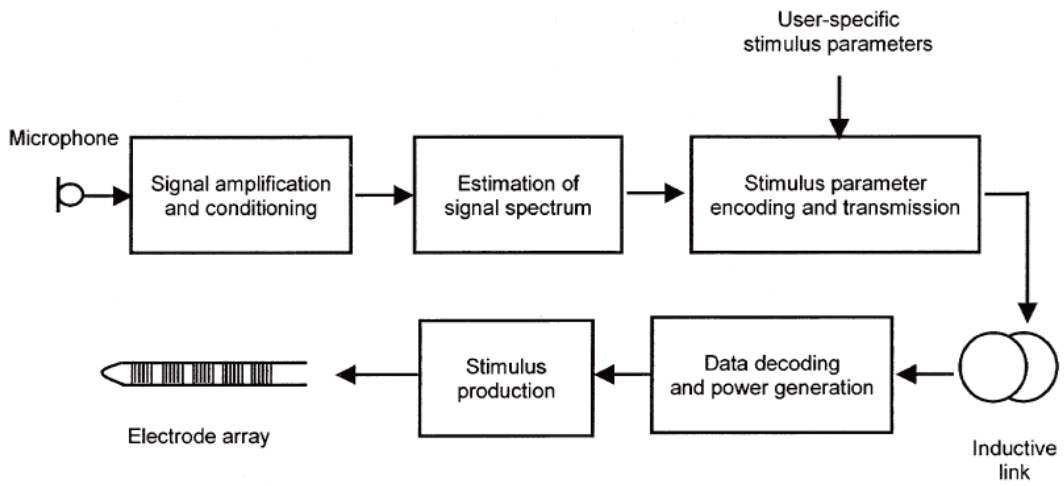

Figure 5: *Diagram of CI functioning, from the perception of the sound by the microphone in the ear's pinna, to the stimulation of the cochlear electrode array. Figure from McDermott (2004)*

A various number of electrodes are mounted on an array implanted into the cochlea. This array can stimulate neurons deeply in the cochlea, corresponding to low-frequencies, but also more basal neurons of the cochlea, corresponding to high-frequencies. There are three main configurations of multiple electrodes that can deliver currents: monopolar, bipolar and common ground (**Figure 6**).

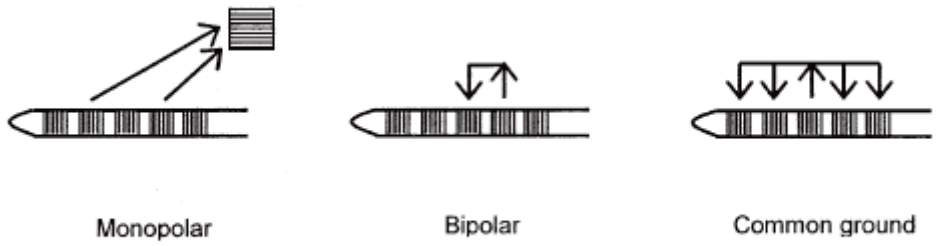

Figure 6*: The three types of electrodes array configuration. Adapted from Mcdermott (2004)*

Monopolar configuration has active electrodes close to the cochlea that flow to one ground electrode, and one or several indifferent electrodes. These latter have larger surface and have a role of current return path for active electrodes. Ideally, in this configuration, active electrodes are close to neural population to excite spatially distinct neurons that allows separation of auditory percepts. In bipolar configuration, both active electrodes are close to neurons and current passes between the two. Various bipolar configurations exist with different benefits depending on the conditions. For instance, active

electrodes can be separated by several inactive electrodes to reduce the threshold current necessary for as stimulus to be audible. Another way to reduce this threshold is to direct the current to the cochlea radially, compared to longitudinal path. This increases the spatial selectivity of electrodes. In the common ground configuration, there is one active electrode and the other electrodes are used for the return of the current.

Sound processors analyze the sound signal and select the appropriate features of this signal to convert it into an electric signal that will be interpreted as hearing. Sound processors have different coding strategies, the principal ones are feature-extracting strategies, spectrum-estimating pulsatile schemes or analog stimulation schemes. 1) Feature-extracting strategies were the first coding approach for CI, but are no longer used for sound processors. This processor extracted the sound signal with an a priori for speech, it extracted fundamental frequency (F0) and frequencies of the two first formants, converting them to electrodes positions. Bandpass filters allowed the estimation of the amplitude of the signal that was transmitted to basal electrodes. Pulsatile stimulation rate was determined by the fundamental frequency of the speech. This feature-extracting strategy was quite efficient for speech recognition. However, when there was background noise, it was really difficult to obtain a satisfactory speech signal as estimation of F0 with different simultaneous sound sources was very difficult. Moreover, as this processor was based on speech signal extraction, non-verbal auditory signals were poorly processed. As a consequence, these strategies for processor were abandoned and new strategies emerged. 2) Spectrum-estimating pulsatile schemes are the most used in CI with multiple-electrodes. These schemes present information extracted from the spectral features of the most prominent sound signals, not necessarily speech. Electrodes of the implant deliver electric pulses that correspond to filters at the sampling rate. Thus, pulse rate is independent of the input signal, and corresponds to thousands pulses per second per channel in most of the implants. Bandpass filters allow to estimate the short-term spectrum of the sound signal, and then envelopes are estimated. Amplitudes of these envelopes are then converted to currents. 3) Analog stimulation schemes are less often use than spectrum-estimating pulsatile schemes. Bandpass filters estimate the short-term spectrum of the sound signals such as in the pulsatile schemes. Rather than using the signal envelopes, it uses the waveforms from each filter. These waveforms are

compressed and delivered as continuously varying currents. In conclusion, numerous sound-processing strategies are available. The most commonly used in the new implants is the spectrum-estimating pulsatile schemes but new strategies emerge every day to allow a better perception of the auditory environment by CI users.

2. *Limitations of the implant*

With growing knowledge of technology, psychophysics and neurosciences, industries were able to design more and more sophisticated implant over the years. There are now over 60000 persons implanted worldwide (McDermott, 2004) and this number increases every year. CI users are now able to understand speech even in the absence of lip reading, by using the device alone. However, auditory perception in cochlear implants (CI) users can remain impaired due to a limited sound frequency discrimination by the implant (Glennon et al., 2020; Lehmann & Paquette, 2015). Indeed, technical constraints of an implant does not allow for a fine-grained decomposition of the auditory signal, compared to the decomposition of a healthy cochlea (McDermott, 2004). As a result, CI users still have difficulties for hearing speech in noise, understanding tone languages, understanding prosody, and music perception, even though there is a high demand of post-lingually deafened CI users to be able to enjoy music again. These non-verbal auditory perception difficulties are assumed to be related to a pitch perception deficit, resulting from the degraded auditory input.

3. *Consequences for non-verbal auditory perception*

See the Introduction or Study 4 for full details about non-verbal auditory deficits in CI users.

In CI users, deficits of music processing affect multiple dimensions, such as pitch processing (Hopyan et al., 2016; Petersen et al., 2020; Sharp et al., 2018) and temporal processing (Jiam & Limb, 2019). These deficits contribute to impaired musical emotion recognition (Ambert-Dahan et al., 2015; Hopyan et al., 2016; Lehmann & Paquette, 2015; Luo & Warner, 2020; Paquette et al., 2018; Shirvani et al., 2014, 2016). Despite these limitations, some CI users still enjoy music, and have great demands to enhance this pleasure of music listening (Fuller et al., 2019).

Even though speech in silent environments is quite well restored by CI users, the pitch deficit is still limiting their non-verbal auditory perception in speech signals (Kalathottukaren et al., 2015). For intentional prosody, CI users have demonstrated poor perception abilities (Lo et al., 2015; Marx et al., 2015; Peng et al., 2008) as well as a deficit for production (Peng et al., 2008). For emotional prosody, CI users have deficits when visual cues are unavailable (Deroche et al., 2019; Pak & Katz, 2019; Paquette et al., 2018). In addition to these difficulties to perceive prosody during a conversation, CI users still experience difficulties to hear and understand speech in noise (Bugannim et al., 2019; Choi et al., 2017; Hong & Turner, 2006).

D. Conclusion

The three conditions previously described in this chapter demonstrate some possibilities of disruptions of the non-verbal auditory pathways. A deficit can be present from the beginning of life (such as CA or early deafness) but can appear at every stage of the life (BD or post-lingual deafness). For CI users, even if technology has emerged to help these patients with incredible advances, numerous challenges are still present. Indeed, participants are still complaining of their poor auditory perception, and therefore of their reduced abilities to communicate with their environment. As technology has its limits, numerous studies are now interested on how improve these participants' life, by using their brain capacities.

III. Training and boosting non-verbal auditory processing in healthy and pathological populations

A. Introduction

Brain plasticity is the capacity of the nervous system to learn and adapt to its environment throughout the lifespan. This mechanism changes the input and the targets of neurons through the modulations of synapses and the number of neurons. These cellular changes in the brain will lead to changes in behaviour and vice versa. Therefore, the brain is continuously changing and adapting to the individual's life. Plasticity in the adult brain is a concept that has been investigated with the findings of adult neurogenesis (Spalding et al., 2013; Toni et al., 2007). However, the term plasticity was first introduced in 1890 by W. James to characterize the possibility of the human being to modify its behaviour. Ràmon y Cajal (1904) then suggested that this modification of behaviour could have an anatomical basis in the brain (Pascual-Leone et al., 2005).

B. Adult plasticity of the auditory system

Here, we will focus on how plasticity is induced and the consequences of these mechanisms at the behavioural level, without describing all the molecular processes occurring during plasticity. A lot of studies about plasticity have been focus on visual processes (see Pascual-Leone et al., 2005 for a review), however here we summarized the research in auditory modality only. Interestingly, plasticity can occur at every steps of the adult life, in healthy situations, but can also be induced after a trauma. This plasticity can be directly linked to one modality but could be enhanced by crossmodal activity (Shimojo & Shams, 2001).

1. Auditory plasticity occurring during the adult life

Originally, the auditory system was thought to be quite rigid with only a possibility of modulation during a short critical period during development. However, recent research has demonstrated that similarly as what has been observed in the visual system, the auditory system is constantly changing and

reorganizing (Irvine, 2018). Plasticity of the auditory system can be driven by bottom-up processes (change of the sensory input), after hearing loss for instance, but also by top-down processes such as attention and learning.

a. Auditory plasticity and learning

Several paradigms have demonstrated rapid changes in the primary AC following learning (Pantev et al., 2015; Pantev & Herholz, 2011). For example, in the tonal conditioning paradigm, a tonal conditioned stimulus leads to an increased response in the AC over training, whereas the pre-training best-frequency response decreases over training and this best-frequency gets closer and closer to the conditioned stimulus (Weinberger, 2004, 2010). As a consequence, the region representing the best-frequency in the AC becomes bigger. This result shows that plasticity following learning can be really quick in the AC and leaves a memory trace. Other paradigms in humans have demonstrated that auditory cortical plasticity can be short or long-term following learning, but mostly depends on the task, the nature of the response in the task and the learning strategy (Irvine, 2018). For instance, musical training have been shown to have effect on AC plasticity in short-term as well as in long-term training (Pantev et al., 2015).

Even if many studies on auditory plasticity have been focused on AC changes, some others have reported that belt areas can also demonstrate changes following learning (Atiani et al., 2014). For instance, in a fear conditioning task in animals, learning was demonstrated to depend on the connection between the AC and the amygdala (Bocchio et al., 2017; Kim & Cho, 2017). Moreover, similar effects of learning on plasticity can be visible also in the medial geniculate body as well as the inferior colliculus (Edeline & Weinberger, 1991; Slee & David, 2015). However, this subcortical plasticity could be partially due to influences from the AC (Suga & Ma, 2003).

b. Auditory plasticity following hearing loss

The first research on auditory plasticity was conducted on hearing loss and the cortical changes associated to this trauma (see Eggermont, 2017 for a complete review). Following a damage to the cochlea, plasticity in the AC occurs. In particular, depending on the extend on the hearing loss, the cortical tonotopic map of the mature AC can change. Generally, the cortex region corresponding to the

area where the damaged region of the cochlea would project, is reorganized and allocated to other frequencies (generally the close frequencies to the damage). Changes following the hearing loss can be seen really quickly in the AC (weeks or months after the trauma) but more long-term changes can also occur. In the adult brain, plastic changes seem to be restricted to the AC and the thalamus only (Irvine, 2010). Even if the auditory plasticity engaged after hearing loss could be seen to enhance the individual's ability to extract the available information in the auditory environment, it appears that it does not compensate for the hearing loss. This auditory plasticity is most likely a homeostatic plasticity that maintains activity in the neuronal circuits, but cannot restore the perception of damaged frequencies (Turrigiano, 2011).

As CI can restore hearing partially, numerous studies have investigated the effect of the implantation on brain plasticity (Glennon et al., 2020). Indeed, following the implantation, remarkable progress can be seen in patients such as in the perception of speech. Numerous studies have investigated auditory plasticity following implantation in animals, such as cats, and showed restoration of normal tonotopic organization even with late implantations (see Fallon et al. (2008) and Kral et al. (2006) for reviews). Moreover, numerous researches are done on congenitally deaf children and the effect of the age of implantation. In sum, the sooner the implantation is done (before the age of 3), the better is the brain response with normal auditory evoked potentials and better language skills (Kral & Sharma, 2012). However, some studies on adult CI users with acquired deafness also demonstrated brain plasticity in these patients. Indeed, the frequency discrimination abilities of CI users was associated with changes of the auditory evoked potentials (shorter latency, increased amplitude) (Sandmann et al., 2015).

2. Crossmodal plasticity

To perceive environmental stimuli, multisensory interactions are essential. For instance, the McGurk effect shows that the integration of visual and auditory information is crucial for speech perception (McGurk & MacDonald, 1976). At the brain level, several evidence has been reported that strong interactions are present across sensory modalities (Shams & Seitz, 2008; Shimojo & Shams, 2001). For example, animal studies involving section of sensory inputs have shown that the thalamic nuclei normally used for auditory signal (medial geniculate nucleus) can be reorganized to welcome retina's

inputs, and directly respond to visual stimulation. In humans, in the case of early sensory deprivation, the cortical area can be reorganized and used by another modality. For example, in deaf individuals, visual event-related potentials tend to be enhanced (Neville et al., 1983; Neville & Lawson, 1987).

Multisensory integration has been reported to be stronger when one of the sensory modalities is deficient (Frassinetti et al., 2005; Grasso et al., 2016; Passamonti et al., 2009). For example, in participants with reduced visual acuity, auditory cues on top of visual cues (and resulting audiovisual interactions) allowed improving their visual detection threshold (Gabor patches) beyond their visual-only performance, which was not observed in control participants (Caclin et al., 2011). Similarly in participants with a pitch processing deficit, such as congenital amusia, visual stimulations improved performance in an auditory pitch task (Albouy, Lévêque, et al., 2015). Numerous studies have investigated audiovisual integration in CI users: first results using speech stimuli and faces showed enhanced auditory capacities in CI users with audiovisual integration (Barone et al., 2016; Rouger et al., 2007, 2008; Strelnikov et al., 2009, 2015). This audiovisual enhancement is increased in CI users compared to normal-hearing participants for all types of speech, even foreign-accented speech (Waddington et al., 2020). One study on non-linguistic speech processing, using voice gender identification, suggests that CI users are influenced more strongly by visual information than are NH participants, even for tasks not directly related to speech comprehension (Barone et al., 2016).

However, crossmodal plasticity can also be a disadvantage in the case when brain regions are not properly reorganized. For instance, in CI users, it has been shown that if the AC was activated by a visual stimulus, this multisensory integration of information could become maladaptive. Indeed, increased coupling between occipital and temporal lobes was associated with poor outcomes of the implantation (Glennon et al., 2020).

C. Training strategies in populations with non-verbal auditory deficit: the example of CI users

1. Introduction

As we have seen in the previous section, CI users are still experiencing numerous challenges regarding the perception of their auditory environment. In particular, non-verbal auditory processing is impaired in these participants with limited hearing in noise or music perception (Carlson, 2020). Moreover, CI users reflect this general limitation in their behaviour and demands (Pattisapu et al., 2020). Cortical plasticity is happening in the brain of CI users at all times, even if the user is particularly old. This gives hope to find non-invasive solution to enhance non-verbal auditory perception in CI users.

Some studies have been interested in the rehabilitation of children with CI (Gfeller, 2016; Rayes et al., 2019; Torppa & Huotilainen, 2019), as they would demonstrate more brain plasticity and therefore potentially more improvements. However, as seen previously, brain plasticity can also occur in adults, all life long, suggesting that training could be efficient in post-lingually CI users too. Here, we review two main training strategies that have been used in link with the auditory brain plasticity previously explained: musical training using purely auditory stimulation and multisensorial training using crossmodal stimulation. Many trainings have been developed to enhance speech perception but here we focus on trainings that specifically improved non-verbal auditory perception in CI users.

2. Musical training

The perception of music requires higher cognitive functions and music processing engages multiple brain regions. Indeed, to process music, an individual need to process the pitch but also temporal and spectral cues, as well as make predictions to perceive the melody and the harmony. Studies comparing musicians and non-musicians' abilities have revealed that musicians can have better abilities in the non-verbal auditory cognition (Talamini et al., 2017). For example, music training could enhance the recognition of emotional prosody (Lima & Castro, 2011; William Forde Thompson et al., 2004). Indeed, some brain research suggested that expertise in one emotional sound category (in this instance, music) could enhance the processing of other emotional sounds such as vocal sounds (Nolden et al., 2017; Patel,

2014). This suggests that music can be a powerful tool to enhance cognition. This is why music training (from simple perception training to instrument training) engages multiple brain regions and requires a high-level cognitive functioning, which can be used for several other processes such as the perception of the auditory scene. As music training provides complex auditory stimulation but also strong interactions between perception and action, it has become more and more common to use music when aiming to rehabilitate participants with hearing deficits such as CI users (Lerousseau et al., 2020).

Several musical trainings have been tested to enhance non-verbal auditory abilities in CI adult users (Barlow et al., 2016; Firestone et al., 2020; Smith et al., 2017) but the most commonly used is the melodic contour identification (MCI) training (Fu & Galvin, 2007). This task consists of 5-note sequences with variating pitch. The pitch range is over 3 octaves and the pitch changes within a sequence can be from 1 to 5 semitones. Subjects have to identify the contour they have heard by selecting the appropriate visual schematic representation (as up and down pattern). Over the past few years, several studies have investigated the effect of this training (with different implementation and duration) in CI users on music perception (Fuller et al., 2018; Galvin et al., 2007), hearing in noise (Lo et al., 2015; Patel, 2014) or intentional (Lo et al., 2015; Patel, 2014) and emotional prosody perception (Fuller et al., 2018). They demonstrated a general improvement of CI users at the MCI task itself and the recognition of familiar melody (Fu & Galvin, 2007; Galvin et al., 2007). Interestingly, a small improvement of hearing in noise was found in some studies (Lo et al., 2015; Patel, 2014) but this effect was not found in all studies (Fuller et al., 2018). An improvement of intentional prosody (recognizing question and statement) was demonstrated following this training in CI users (Lo et al., 2015; Patel, 2014), as well as a better recognition of emotional prosody (Fuller et al., 2018). However, these results remain very little and as no improvement of the quality of life of CI users was reported (Fuller et al., 2018).

3. Training using multisensorial integration

As multisensory stimulation is what humans are more often facing in their everyday life, it was suggested that multisensory training would be closer to a natural environment, hence leading to more efficient learning (Shams & Seitz, 2008). Indeed, even for unisensory learning, multisensory integration

can facilitate learning. For example, in a voice recognition task, people who had audiovisual training (with video of the speaking face), had better scores than compared to audio training only. In this process of learning, congruency between visual and audition seems to play an essential role. Indeed, it could help the learning, based on previous knowledge of the individual. The mechanisms of multisensory learning recruit large brain networks. The main pathway involves that neurons of one modality receive signal from neurons of other modalities. In the process of learning, neurons need to pass a firing threshold (in term of rate, oscillation, response latency), input from other modality neurons could help to pass this threshold (Shams & Seitz, 2008).

Animal studies have revealed that multisensory audiovisual training improves performance in cochlear implanted ferrets or cats with hearing loss. This training performance was correlated with the degree of primary AC responsiveness (Glennon et al., 2020). These results suggest that crossmodal training strategies could be efficient in adults with CI. Two main strategies have been tested so far, one using the coupling of auditory and motor processing, another one using audiovisual coupling.

a. Auditory-motor/tactile training

Recently, researchers have been interested in the coupling between auditory and motor systems to enhance the perception of non-verbal auditory cues in CI users (Chari et al., 2020; Huang, Chang, et al., 2017; Huang et al., 2020; Huang, Sheffield, et al., 2017). Based on musical training previously described, one study looked at the effect of a contour auditory-motor training in CI users (Chari et al., 2020). CI users had to train to reproduce the contour of a melody they heard on a keyboard. CI users only improved at the MCI task following this training but no improvement of hearing in noise or prosody perception was found. Similarly, another research group was interested at the potential use of electrotactile stimulation in combination with CI to enhance their perception of the auditory scene (Huang, Chang, et al., 2017; Huang et al., 2020; Huang, Sheffield, et al., 2017). In sum, electrotactile stimulation transforming the fundamental frequency of the sound into tactile vibration could enhance non-verbal auditory perception in CI users. In particular, CI users had enhanced tone and speech recognition in

noise (Huang, Chang, et al., 2017; Huang, Sheffield, et al., 2017) as well as a better melody recognition (Huang et al., 2020).

However, these trainings are still very limited in time and in number of CI users included in the studies. They need further replications and long-term developments to be confirmed.

b. Audiovisual training

Based on previous results in CI users (Innes-Brown et al., 2011; Rouger et al., 2007, 2008; Strelnikov et al., 2015), it seems that audiovisual training could be a good strategy to enhance auditory perception in CI users. However, only few studies have evaluated the potential benefit of audiovisual training on CI users (Bernstein et al., 2014; Sato et al., 2019; Vandali et al., 2015). Two studies used audiovisual speech and looked at the effect of this training in CI users on their speech perception in noise (Bernstein et al., 2014; Sato et al., 2019). The both used spoken words in association with visual stimuli relevant (faces) or irrelevant (pictures associated to words). They demonstrated an enhanced perception of speech in noise with this audiovisual training, at least at the end of the training. One study used matching pitch/timbre and visual patterns to train CI users and looked at their frequency discrimination scores (Vandali et al., 2015). They demonstrated an enhanced F0 discrimination in CI users after the training but only if the task was unimodal and not if the task combined multiple cues (pitch, timbre), making it difficult to extend to music perception abilities.

In conclusion, only very few studies have been done to evaluate the potential benefit of multisensory trainings in CI users for non-verbal auditory perception. Even if the musical training studies show promising results, the actual knowledge about multisensory integration in CI users let us hope for even better results with audiovisual trainings.

Objectives

Non-verbal auditory cues become essential in many situations when auditory conditions are not optimal for speech comprehension. For instance, in everyday life we face many noisy situations: from a phone conversation to environmental noise. In these situations, non-verbal auditory processes such as auditory streaming to separate the source of interest from the noise, but also emotional prosody to detect the intentions of the speaker, become essential to communicate. Numerous deficits can deteriorate the perception of non-verbal auditory cues. From central auditory disorder such as congenital amusia or brain-damaged patients to peripheral disorder such as hearing loss, multiple deficits are associated with difficulties to process non-verbal auditory cognition.

The first axis of this PhD was the characterization of non-verbal auditory cognition in several deficits. First of all, in congenital amusia, even if the deficit was first described as music-specific, it appears that speech-related processing could be impacted. Moreover, as congenital amusia was characterized as a pitch deficit, and pitch is essential for emotional prosody perception, we wondered if emotional prosody was impacted in congenital amusia. To do so, we designed two experiments to test this hypothesis. Study 1 tested behaviorally the potential deficit of emotional prosody recognition in congenital amusia. Based on this study, we found a dichotomy between explicit and implicit perception of emotional prosody in congenital amusia and tested this hypothesis with electrophysiology (Study 2). Similarly, we wanted to better characterize the emotion perception pattern in brain-damaged patients. As many studies were done using faces and prosody material, we wondered about their musical emotion perception. We tested this using a similar protocol as in congenital amusia with music emotional stimuli (Study 3).

In the second axis of this PhD, we aimed to set up a new rehabilitation program for non-verbal auditory perception in case of a deficit. As CI users are in high demand for this rehabilitation program, we focused our research on this population. In the first study (Study 4), we designed a new testing battery to assess non-verbal auditory perception deficits in CI users. In the second study, we designed a new rehabilitation program based on audiovisual interaction. We first tested this program on healthy participants to validate the feasibility and plan to test it with CI users in the future. Here, we report only the preliminary results of three pilot participants as the experiments were stopped due to the confinement in France.

Altogether, this PhD work aim to provide better understanding of non-verbal auditory cognition in several deficits and brings new perspectives for a new rehabilitation program for these deficits.

Empirical part

Part 1: Emotion perception in non-verbal auditory deficits

IV. Emotional prosody perception in congenital amusia

A. Objectives

While congenital amusia was first described as a music-specific deficit, recent studies have shown that this deficit could extend to the speech domain (Liu et al., 2015; Patel et al., 2008; Tillmann, Rusconi, et al., 2011). Moreover, as previous studies demonstrated some deficit of congenital amusics for emotion perception (Lévêque et al., 2018; Lima et al., 2016; Lolli et al., 2015; Thompson et al., 2012), we were interested in characterizing the perception of emotional prosody in congenital amusia. In the first study, we used a behavioral paradigm to determine if congenital amusics had difficulties perceiving emotional prosody, and if so, which acoustic cues could be the origin of this deficit. As congenital amusia was recently suggested to be a deficit of consciousness rather than a deficit of perception *per se*, we used a behavioral paradigm with two tasks: explicit emotion recognition and implicit intensity ratings of these emotions.

Based on the results of the first study, we suggested that congenital amusics could have a deficit of explicit recognition of emotional prosody, especially for short vowels. But this deficit was not visible for implicit intensity ratings. To further test this hypothesis, we designed an oddball paradigm using electroencephalographical measures and using the same vowel stimuli as in the behavioral study. In this second study, we try to understand the amusics' brain response to emotional prosody and emotionally neutral stimuli in pre-attentive context.

B. Study 1: Emotional prosody in congenital amusia: Impaired and spared processes

Neuropsychologia 134 (2019) 107234

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Emotional prosody in congenital amusia: Impaired and spared processes

A. Pralus [a,b,*], L. Fornoni [a,b], R. Bouet [a,b], M. Gomot [c], A. Bhatara [d], B. Tillmann [a,b,1], A. Caclin [a,b,1]

[a] Lyon Neuroscience Research Center, CNRS, UMR5292, INSERM, U1028, Lyon, F-69000, France
[b] University Lyon 1, Lyon, F-69000, France
[c] UMR 1253, INSERM, Université de Tours, Tours, France
[d] UMR 8242, Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France

ARTICLE INFO

Keywords:
Tone deafness
Emotion
Language
Music
Explicit and implicit processes

ABSTRACT

Congenital amusia is a lifelong deficit of music processing, in particular of pitch processing. Most research investigating this neurodevelopmental disorder has focused on music perception, but pitch also has a critical role for intentional and emotional prosody in speech. Two previous studies investigating amusics' emotional prosody recognition have shown either some deficit or no deficit (compared to controls). However, these previous studies have used only long sentence stimuli, which allow for limited control over acoustic content. Here, we tested amusic individuals for emotional prosody perception in sentences and vowels. For each type of material, participants performed an emotion categorization task, followed by intensity ratings of the recognized emotion. Compared to controls, amusic individuals had similar recognition of emotion in sentences, but poorer performance in vowels, especially when distinguishing sad and neutral stimuli. These lower performances in amusics were linked with difficulties in processing pitch and spectro-temporal parameters of the vowel stimuli. For emotion intensity, neither sentence nor vowel ratings differed between participant groups, suggesting preserved implicit processing of emotional prosody in amusia. These findings can be integrated into previous data showing preserved implicit processing of pitch and emotion in amusia alongside deficits in explicit recognition tasks. They are thus further supporting the hypothesis of impaired conscious analysis of pitch and timbre in this neurodevelopmental disorder.

1. Introduction

Congenital amusia is a lifelong deficit of music perception and production, also referred to as "tone deafness". This disorder is estimated to affect one to four percent of the general population (Peretz et al., 2007; Peretz and Vuvan, 2017) and is suggested to have genetic origins (Peretz et al., 2007). Amusic individuals have neither peripheral auditory deficits nor brain lesions, but they are unable to detect out-of-key notes in a melody, and they sing out-of-tune (Peretz, 2016; Tillmann et al., 2015). Pitch processing deficits were observed for amusic individuals in perception tasks, such as pitch discrimination (Hyde and Peretz, 2004) or pitch contour change detection (Peretz et al., 2003), as well as in short-term memory related to pitch (Tillmann et al., 2016b). In a typical (non-amusic) brain, a fronto-temporal network is involved in pitch processing and memory, and thus also in music perception (Gaab et al., 2003; Koelsch et al., 2009; Zatorre et al., 1994). In the amusic brain, anatomical and functional abnormalities have been observed in this fronto-temporal network (Albouy et al., 2013a; Hyde et al., 2007; Hyde et al., 2006; Hyde et al., 2011). More specifically, decreased fronto-temporal connectivity was observed in congenital amusia, in particular in the right hemisphere, together with an increased connectivity between the auditory cortices (Albouy et al., 2013a; Albouy et al., 2015; Hyde et al., 2011; Leveque et al., 2016; Loui et al., 2009; Tillmann et al., 2015). These findings suggest an altered auditory neural network underlying the pitch processing deficit in congenital amusia (Leveque et al., 2016). The deficit extends to timbre (Stewart, 2011; Tillmann et al., 2009), whereas temporal processing seems to be mostly preserved in amusia (Hyde and Peretz, 2004), at least when the material does not entail pitch variations (Foxton et al., 2006; Pfeuty and Peretz, 2010).

Given these processing deficits for music, emotion processing has

Abbreviations: MBEA, Montreal Battery for the Evaluation of Amusia; PDT, Pitch Discrimination Threshold.

* Corresponding author. CRNL, 95 Bd Pinel, 69675, Bron Cedex, France.
E-mail address: agathe.pralus@inserm.fr (A. Pralus).

[1] Equally contributing authors.

https://doi.org/10.1016/j.neuropsychologia.2019.107234
Received 31 July 2018; Received in revised form 12 August 2019; Accepted 16 October 2019
Available online 21 October 2019
0028-3932/© 2019 Elsevier Ltd. All rights reserved.

also been studied with musical material in congenital amusia. Despite impaired perception and memory of music, some listeners afflicted with congenital amusia have been reported to either like or avoid listening to music. This dichotomy occurs independently of the severity of amusia as measured by the Montreal Battery for the Evaluation of Amusia (McDonald and Stewart, 2008; Omigie et al., 2012). These subjective reports about liking/avoidance inspired recent studies investigating musical emotion processing in congenital amusia. Gosselin et al. (2015) showed no impairment of emotion recognition (Gosselin et al., 2015), while Leveque et al. found a mild impairment (Lévêque et al., 2018). Similarly, a study that focused on dissonance/consonance judgments of musical materials reported that congenital amusics were able to recognize the suggested musical emotions, but they based their judgments more on roughness rather than on the harmonicity cues used by control participants (Marin et al., 2015). These findings and previous reports of the perceptual deficits in amusia suggest that amusics' emotional judgments in music are based largely on roughness and tempo rather than harmonicity and mode cues (Gosselin et al., 2015; Lévêque et al., 2018).

As pitch processing is involved not only in music processing, but also in speech processing, several studies have focused on speech perception abilities in amusia. Interestingly, while early studies did not report deficits in speech processing (Ayotte et al., 2002) or in memory for verbal materials (Tillmann et al., 2009; Williamson and Stewart, 2010), subsequent studies using more fine-grained materials and methods did reveal speech processing impairments in amusia. Specifically, intonation recognition and perception of speech contour is impaired across languages - this includes tonal languages, non-tonal languages, and even artificial verbal materials (Jiang et al., 2010, 2012; Liu et al., 2017; Liu et al., 2015; Nan et al., 2016; Nguyen et al., 2009; Tillmann et al., 2011a, 2011b). In speech, prosody is essential to detect a speaker's intentions and emotions, that is, to understand what the speaker means and to follow and participate in a conversation. Pitch changes related to prosody can be of rather large extent, and other acoustic cues (e.g., loudness, duration) also contribute to prosody. The cerebral correlates of prosody processing involve bilateral inferior frontal gyri (Frühholz et al., 2012). In addition, intentional prosody processing also involves the middle superior temporal gyrus, whereas emotional prosody processing involves the right anterior superior temporal gyrus (Frühholz et al., 2012; Liebenthal et al., 2016). To decode the emotional meaning of speech, the amygdala, the insula and the auditory cortex are also involved (Frühholz et al., 2016). These regions are also more generally involved in decoding emotional information, whether it occurs in music or speech materials (Bestelmeyer et al., 2014; Escoffier et al., 2013). As the right superior temporal and inferior frontal regions participate in emotional prosody processing, and these regions exhibit differences in amusia compared to controls (Albouy et al., 2013a; Hyde et al., 2007), Liu et al. (2015) suggested that differential processing in these regions could underlie amusics' difficulties processing subtle emotional and prosodic changes (Ayotte et al., 2002; Patel et al., 2005). Thompson et al. (2012) showed a deficit in the recognition of emotional prosody in amusic individuals, especially for happiness, tenderness, irritation and sadness. They suggested that this deficit might be due to the impaired processing of acoustic features that are relevant for emotional information in music and speech, such as, for example, pitch. In contrast, Lolli et al. (2015) did not find any differences in emotion recognition in speech between amusics and controls. However, amusic individuals were impaired when speech was low-pass filtered, suggesting that low frequencies in speech carry emotional content and that amusics are less sensitive to these cues. Thompson et al. and Lolli et al. used spoken sentences of the Macquarie Battery of Emotional Prosody (MBEP; the same 84 sentences were used in both studies), which were read by actors in a natural way. In this material, numerous acoustic parameters vary depending on the emotion conveyed, so amusic participants could infer emotions using non-pitch cues, such as the intensity or the duration of vowels, words, or sentences. In addition, the differences between the results of these two studies could be due to differences in amusic recruitment criteria. Lolli et al. recruited amusics based on their pitch discrimination threshold, whereas Thompson et al. recruited amusics based on the three MBEA pitch subtests (scale, interval, contour). To further explore the potential auditory emotion processing deficit in amusia, Lima et al. (2016) investigated amusics' emotion recognition not only in speech, but also in non-verbal auditory stimuli. Participants heard non-verbal vocalizations (e.g. screams, sobs, sighs) that were not directly related to speech and were asked to judge the degree of seven expressed emotions on multiple seven-point scales. Results revealed that response accuracy was decreased in congenital amusic individuals relative to the control participants for both verbal and non-verbal stimuli. However, the task involving multidimensional ratings requires making multiple comparisons, so it also involves memory. Indeed, participants had to rate the intensity of each of seven emotions sequentially, which might thus involve a rather strong memory component to do the task. Some studies have suggested that amusia could be due to a deficit in pitch memory rather than in pitch discrimination (Albouy et al., 2016; Albouy et al., 2015; Albouy et al., 2013a, 2013b; Tillmann et al., 2016b; Tillmann et al., 2009). A memory deficit such as this could have influenced Lima et al.'s results. Lima et al. also investigated the role of acoustic parameters (F0, intensity and spectral centre of gravity) in emotion recognition and showed that the combinations of acoustic cues used to determine prosodic emotions differed between amusics and controls. However, the variation in the duration of the stimuli and their rhythmic structure (specifically, laughter vs. other stimuli) across emotions could have provided a variable amount of acoustic cues in the stimuli depending on the emotion, making it difficult to pinpoint the impairment for emotion processing in amusia. The conflicting results of previous research on emotional prosody processing in congenital amusia led us in the present study to use not only sentences, but also vowels, which allowed us to reduce the variation of acoustic parameters within the stimuli. In isolated vowel sounds, temporal cues (such as duration or syllable rate differences) can be controlled, whereas this is not possible in full sentences. This material choice allowed us also to further investigate the acoustic parameters that could influence emotional prosody recognition and whose processing could be impaired in congenital amusia, beyond the pitch deficit (Marin et al., 2015; Tillmann et al., 2009; Whiteford and Oxenham, 2017; Zhang et al., 2017). In addition, we used a paradigm including two judgments of the verbal stimuli; participants were first asked to recognize the emotion with a forced-choice task, and then to judge the intensity of the recognized emotion. This paradigm allowed us to evaluate whether the deficit of amusia is in recognizing an emotion (as measured with the categorization task) or occurs as part of a more general evaluation of an emotion (as measured with the intensity ratings).

Previous studies investigating emotions in music and speech materials in congenital amusia used mostly explicit emotion recognition tasks, requiring participants to choose among given categories. However, indirect investigation methods have revealed preserved implicit processing of pitch and music in amusia alongside disrupted explicit processing (Omigie et al., 2013; Tillmann et al., 2014; Tillmann et al., 2012; Tillmann et al., 2016a). For example, amusic individuals process pitch changes or pitch incongruity (Peretz et al., 2009; Zendel et al., 2015), even if they cannot explicitly detect it (Moreau et al., 2009; Tillmann et al., 2016a). Based on these observations, Leveque et al. also asked participants to rate the intensity of the emotions they previously recognized in musical excerpts, as a potential measure of implicit knowledge of musical emotions (Lévêque et al., 2018). Indeed, intensity ratings of emotions can be done with less precise internal representations of these emotions. No verbal or categorical representation of the emotion is necessary; a global appreciation of the stimulus suffices (Lévêque et al., 2018). Interestingly, normal intensity ratings were observed in amusics compared to controls, suggesting that implicit processing of music-induced emotions is preserved in amusia (Lévêque et al., 2018). More generally, recent studies on congenital amusia

suggest that this developmental deficit might be a disorder of consciousness and/or conscious access to pitch representations (Albouy et al., 2016; Marin et al., 2015; Moreau et al., 2009; Moreau et al., 2013; Omigie et al., 2013; Peretz, 2016; Stewart, 2011; Tillmann et al., 2016a). Therefore, in our study, we also asked participants to rate the intensity of the just-recognized emotion as this measure could reflect implicit processing of emotional prosody in congenital amusia.

In sum, emotion perception in amusia is still understudied, in particular for speech, with conflicting results for full sentence materials. Individuals' capacity to process prosody might depend on the type of speech stimuli used; for example, in sentences and non-verbal emotional stimuli, the amount and types of temporal cues might differ between emotions, which should not be the case for short vowel stimuli. Prosody-processing capacity might also depend on the type of task used, specifically whether it is explicit or implicit. Our present study tested the perception of emotional prosody in congenital amusics and matched controls with sentences and short vowels using both an explicit categorization task and intensity ratings (to access a more implicit processing of emotions). While emotional prosody perception has been previously investigated in amusia, it was mostly done with long sentence materials. In contrast, our present study also used short stimuli for this investigation, aiming to detect more fine-grained deficits in amusics and determine related acoustic features. We hypothesized that stronger deficits are observed for emotion categorization in amusics with vowel stimuli than with sentence stimuli. Given that single vowels convey fewer temporal cues, listeners are forced to rely on pitch and other fine spectral and spectro-temporal cues to recognize the emotional content. In addition, numerous studies have demonstrated the importance of using explicit *and* implicit measures to fully understand amusics' perceptual capacities (Albouy et al., 2016; Lévêque et al., 2018; Marin et al., 2015; Moreau et al., 2013; Stewart, 2011; Tillmann et al., 2016a). Given the recent findings of preserved emotional intensity judgements with *musical* stimuli in congenital amusia (Lévêque et al., 2018), we hypothesize that intensity ratings would also be preserved for *speech* stimuli, both for sentences and vowels.

2. Methods

2.1. Participants

Eighteen amusic participants and eighteen control participants matched for gender, age, laterality, education, and musical training were included in the study (Table 1). They all gave written informed consent to participate in the experiment. Prior to the main experiment, all participants were tested with a subjective audiometry, the Montreal Battery of Evaluation of Amusia (Peretz et al., 2003) to diagnose amusia, and a Pitch Discrimination Threshold (PDT) test (Tillmann et al., 2009). A participant was considered amusic if he/she had a global MBEA score below 23 (maximum score = 30) and/or a MBEA pitch score (average of the first three subtests of the MBEA) inferior to 22 (maximum score = 30). All control participants had a global MBEA score higher than 24.5 and a MBEA pitch score higher than 23.3 (see Table 1). All participants had normal hearing (hearing loss inferior to 30 dB at any frequency in both ears). Study procedures were approved by a national ethics committee and participants were paid for their participation.

Prior to the study, 15 amusic and 15 control participants (among the cohort previously described) filled out a questionnaire about their relationships to music and language in their everyday life (based on questionnaires of McDonald and Stewart, 2008, Sloboda et al. (2005)), and Peretz et al. (2009)). Among 90 questions, we selected 14 questions about musical emotions with 9 positive (e.g., "Music can relax me or calm me in a period of stress") and 5 negative ("I never had a chill while listening to music") statements. Participants rated their agreement to these statements from 1 (totally disagree) to 5 (totally agree). For each participant, an average musical emotions score was calculated (with a reverse coding for negative agreements). We also selected 7 questions about speech perception, excluding simple understanding of a language with 6 positive (e.g "Can you tell if someone has an accent?") and one negative ("Do you have difficulties recognizing environmental sounds?") questions. Participants rated to what extent these situations apply to them from 1 (almost every time) to 4 (almost never). For each participant, an average language comprehension score was calculated (with a reverse coding for the negative question; see Table 1). Both the musical emotions score ($t(28) = 2.38$, $p = 0.032$) and the language perception score ($t(28) = 2.29$, $p = 0.038$) differed significantly between the amusics and the controls, with amusics having a lower musical emotions score (i.e. reduced emotional effect of music) and a higher language comprehension score (i.e. weaker language comprehension).

Table 1
Characteristics of the participants in both groups. The MBEA (Montreal Battery for the Evaluation of Amusia, Peretz et al., 2003) score corresponds to the average of the six subtests of the battery (maximum score = 30, cut off: 23). Pitch mean score corresponds to the average of the three pitch subtests in the MBEA (scale, contour and interval, cut off: 22). PDT: Pitch Discrimination Threshold (see Tillmann et al., 2009). Musical emotion score corresponds to the average response to the 14 questions about musical emotion in the questionnaire (high score corresponds to high emotional reaction to music). Speech perception score corresponds to the average response to the 7 questions about speech comprehension in the questionnaire (low score corresponds to a good speech comprehension). The standard deviation is indicated in parentheses. Groups were compared with t.tests (two sided), except for handedness where a Chi2 test was used (Qobs = 0.36).

	Amusics (n = 18)	Controls (n = 18)	p-value (group comparison)
Age (years)	32.67 (±14.88)	35.27 (±14.83)	0.60
Education (years)	15.17 (±2.48)	15.89 (±1.78)	0.32
Musical education (years)	0.056 (±0.24)	0	0.33
Sex	11F, 7M	11F, 7M	
Handedness	2L, 16R	1L, 17R	0.55
MBEA score	21.55 (±1.80) Min: 16.83 Max: 24.17	26.60 (±1.09) Min: 24.5 Max: 28.5	<0.001
MBEA pitch mean score	20.67 (±2.08) Min: 15.67 Max: 23.67	26.78 (±1.53) Min: 23.33 Max: 29	<0.001
PDT (semitones)	1.72 (±1.46) Min: 0.33 Max: 4.99	0.22 (±0.13) Min: 0.08 Max: 0.54	<0.001
Musical emotion score (N = 15 in both groups)	3.85 (±0.87) Min: 2.21 Max: 4.93	4.41 (±0.35) Min: 3.86 Max: 5	0.032
Speech perception score (N = 15 in both groups)	2.44 (±0.35) Min: 1.86 Max: 3.14	2.13 (±0.30) Min: 1.71 Max: 2.71	0.038

2.2. Stimuli

Sentences. Twenty sentences were selected from a larger material set of semantically neutral sentences uttered with different emotions by male and female actors. This selection was based on a pretest by a cohort of non-musician participants (N = 16) in an explicit recognition task. The selected sentences were chosen according to their high recognition scores of the intended emotions (all superior to 80%, mean ± SD = 91.2% ± 7.5). For each emotion (joy, neutral, sadness, anger, fear), four sentences were selected, half pronounced by a male voice and half by a female voice. For each emotion and gender, there were two sentences in French: "J'espère qu'il va m'appeler bientôt" ("I

hope he will call me soon"), and "L'avion est presque plein" ("The plane is almost full"). Stimuli lasted on average 1470 ms (±278 ms) and were equalized in Root Mean Square (RMS) amplitude.

Vowels. Twenty productions of the vowel/a/were selected from a larger material set, all produced with female voices[2] (Charpentier et al., 2018). All stimuli lasted 400 ms and were equalized in RMS amplitude. This selection was based on a pretest by the same control participants as for the sentence material (N = 16). Four stimuli for four basic emotions (joy, sadness, anger, fear) and neutrality were chosen according to their high recognition scores (all superior to 69%, mean ± SD = 82.9% ± 7.4).

Acoustic parameters (pitch mean, pitch slope, spectral flux mean, spectral flux slope, spectral spread mean, spectral spread slope, brightness mean, brightness slope, roughness mean, roughness slope, spectral centroid mean, spectral centroid slope, inharmonicity mean, inharmonicity slope, and attack time) of the stimuli were computed with the MIR toolbox (Lartillot and Toiviainen, 2007; Fig. 1). Recent findings have suggested that amusics' deficit extends beyond pitch perception, notably to timbre perception (Marin et al., 2015; Tillmann et al., 2009; Whiteford and Oxenham, 2017; Zhang et al., 2017). As timbre can be determined by various spectral, temporal and spectro-temporal features, the exact features being still in a matter of debate (Allen et al., 2018; Caclin et al., 2005), we chose to measure several spectral (pitch mean, spectral spread mean, brightness mean, roughness mean, spectral centroid mean), temporal (inharmonicity mean, attack time) and spectro-temporal (pitch slope, spectral flux mean, spectral flux slope, spectral spread slope, brightness slope, roughness slope, spectral centroid slope, inharmonicity slope), parameters in the stimuli. Each parameter (except Attack Time) was computed with a temporal frame of 50 ms by default. We then computed the average of each parameter across time, and the slope of variation of these parameters across time. For sentences, the duration of the stimuli was also computed. For sentences and for vowels, an Analysis of Variance (ANOVA) was performed on each acoustic parameter with emotion (joy, sadness, anger, fear, neutral) as a between-stimuli factor. For sentences, there was a significant effect of emotion for pitch mean ($F(4,15) = 3.14$, $p = 0.046$), brightness mean ($F(4,15) = 3.2$, $p = 0.044$), spectral flux mean ($F(4,15) = 5.64$, $p = 0.006$), spectral flux slope ($F(4,15) = 6.47$, $p = 0.003$), spectral spread mean ($F(4,15) = 6.07$, $p = 0.004$), spectral spread slope ($F(4,15) = 6.62$, $p = 0.003$), roughness slope ($F(4,15) = 3.38$, $p = 0.037$), and spectral centroid mean ($F(4,15) = 4.14$, $p = 0.019$.). After Bonferroni correction for multiple comparisons, only spectral flux slope and spectral spread slope had a significant effect of emotion ($p = 0.003$). According to Fisher LSD post-hoc tests, Joy had a significantly higher pitch than Neutrality ($p = 0.024$). Joy was brighter compared to Neutrality ($p = 0.04$). Joy and Fear had higher mean spectral flux than Sadness and Neutrality (all $p < 0.05$). Anger had stronger slope of spectral flux than Sadness and Neutrality (all $p < 0.04$). Joy, Fear and Anger had significantly lower spectral spread mean than Neutrality (all $p < 0.05$). Anger had less steep spectral spread slope than Sadness and Neutrality (all $p < 0.02$). Anger had significantly stronger roughness slope than Neutrality ($p = 0.04$). Fear had significantly lower spectral centroid mean than Sadness ($p = 0.03$).

For vowels, there was a significant effect of emotion for mean pitch ($F(4,15) = 23.24$, $p < 0.001$), mean brightness ($F(4,15) = 5.85$, $p = 0.005$), brightness slope ($F(4,15) = 4.32$, $p = 0.016$), mean spectral flux ($F(4,15) = 36.52$, $p < 0.001$), roughness slope ($F(4,15) = 3.2$, $p = 0.043$), spectral centroid slope ($F(4,15) = 6.29$, $p = 0.004$), and mean inharmonicity ($F(4,15) = 28.74$, $p < 0.001$). After Bonferroni correction for multiple comparisons, only mean pitch, mean spectral flux and mean inharmonicity had a significant effect of emotion ($p < 0.001$). Fear had a significantly higher pitch than the other emotions (all $p < 0.01$) (as also reported by Lima et al. (Lima et al., 2016, Table 4) and Thompson et al. (Thompson et al., 2012, Table 3) for their materials), and Joy had a significantly higher pitch than Neutrality and Sadness (all $p < 0.004$). Joy was significantly brighter than Anger, Sadness and Neutrality (all $p < 0.05$), and Fear was significantly brighter than Sadness ($p = 0.05$). Anger had a significantly less steep brightness slope than Joy and Neutrality (all $p < 0.039$). Joy, Anger and Fear had significantly higher mean spectral flux than Sadness and Neutrality (all $p < 0.001$), and Sadness had a higher mean spectral flux than Neutrality ($p = 0.05$). Fear had significantly stronger roughness slope than Sadness ($p = 0.037$). Joy and Neutrality had significantly stronger spectral centroid slope than Sadness and Anger (all $p < 0.05$). Joy and Anger had significantly higher inharmonicity means than Fear, Sadness and Neutrality (all $p < 0.01$), and Fear had a significantly higher inharmonicity mean than Neutrality ($p = 0.003$).

2.3. Procedure

The experiment took place in a sound-attenuated room. In each trial, participants listened to a stimulus and were then asked to select the recognized emotion from five options (joy, neutral, sadness, anger, fear). After having given their response, they were asked to rate the intensity of the selected emotion from 1 (not intense) to 5 (very intense), except for stimuli judged as neutral. After the intensity rating response, the following stimulus was automatically played after a variable delay of 1250 ms on average (ranging from 1000 to 1500 ms). The stimuli were presented in two blocks: sentences in one and vowels in another. The presentation order of the two blocks was counter-balanced across participants. Within a block, the presentation order of the stimuli was randomized for each participant, with the constraint that a given emotion cannot be presented more than three times in a row. Presentation software (Neurobehavioral systems, Albany, CA, USA) was used to present the stimuli to the participants and to record responses on a keyboard. The duration of the experiment was 15 min.

2.4. Data analyses

For each participant and emotion, the percentage of correct responses was calculated, separately for sentences and vowels, and for correctly categorized trials, the average ratings of intensity were calculated, again separately for sentences and vowels. In light of recent criticism of classical frequentist analysis (Lee and Wagenmakers, 2014; Wagenmakers et al., 2018), a Bayesian approach was used to analyze the data. Bayesian analyses allow to perform model comparison and to select the best model (with the best evidence), given the data. We analyzed the data with the software JASP (Wagenmakers et al., 2017), with a 2 × 5 Bayesian mixed repeated-measures analysis of variance (ANOVA) on percentages of correct responses and intensity ratings with group (amusic or control) as a between-participants factor, and emotion (joy, sadness, fear, anger or neutral) as a within-participant factor (note that for intensity rating, the emotion factor had only four levels as neutral stimuli were not rated for intensity). We also analyzed the data of sentences and vowels together with group (amusic or control) as between-participants factor, and emotion (joy, sadness, fear, anger or neutral, the latter only for percentage of correct responses) and material (sentences or vowels) as within-participant factors. We reported Bayes Factor (BF) as a relative measure of evidence. To interpret the strength of evidence, we considered a BF under three as weak evidence, a BF between three and 10 as positive evidence, a BF between 10 and 100 as strong evidence and a BF higher than 100 as a decisive evidence (according to (Lee and Wagenmakers, 2014). BF_{10} indicates the evidence of H1 compared to H0, $BF_{inclusion}$ indicates the evidence of one effect over all models. As no post-hoc tests with correction for multiple comparison have as yet been developed for Bayesian statistics (Wagenmakers et al., 2017), we used t-tests with Bonferroni correction for multiple

[2] The vowel stimuli were intended to be also useable in follow-up EEG experiments (with the aim to record the MMN), where it is advisable to use only voices of similar F0, which was the reason why only female voices were recorded.

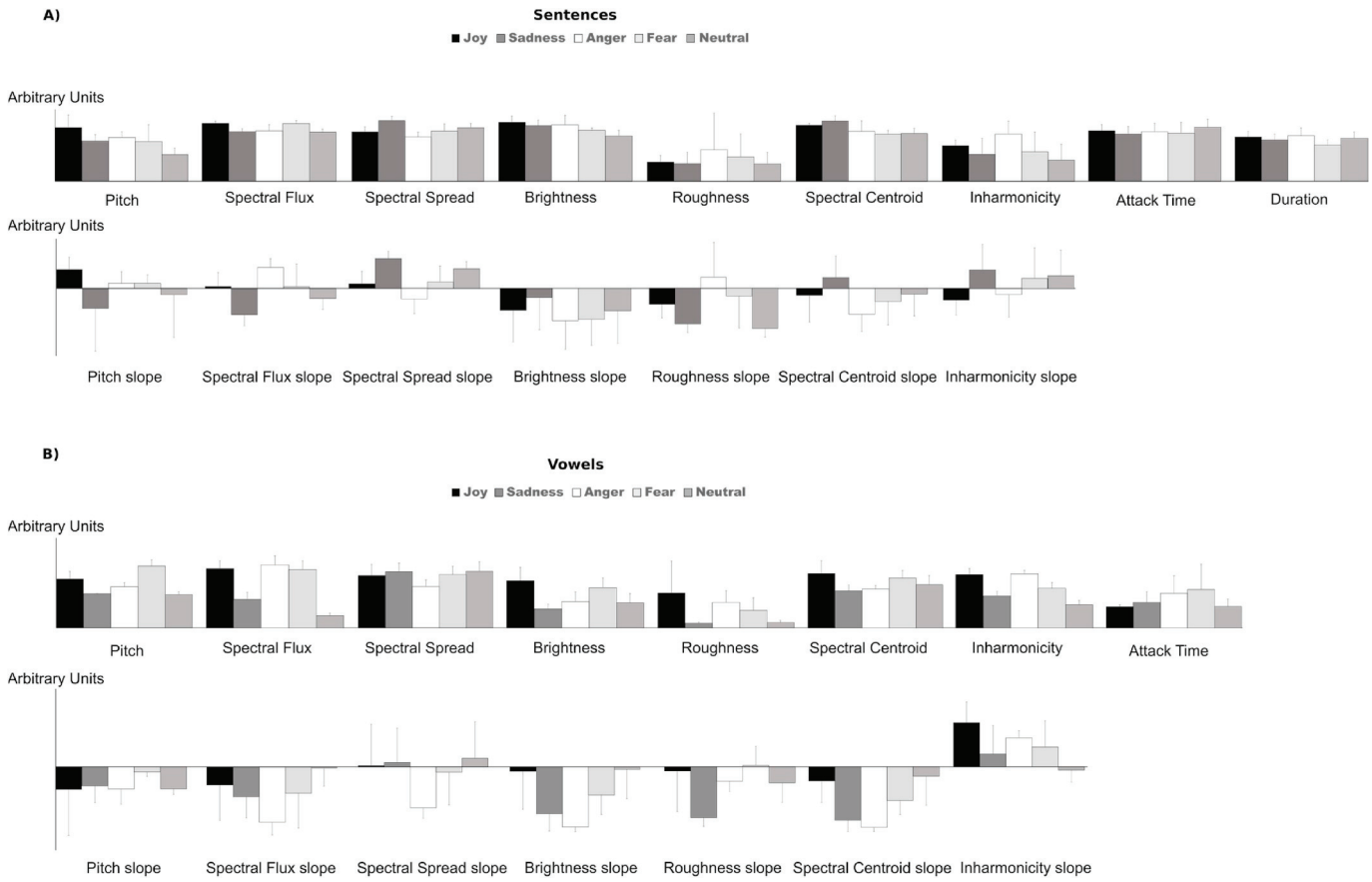

Fig. 1. Acoustic parameters of the stimuli, averaged across all exemplars of a given emotion category. A) Parameters for the sentence stimuli. The first line represents the mean of the parameters (across time), the second line the slopes of their variation across time. There was a significant effect of emotion for pitch mean, spectral flux mean and slope, spectral spread mean and slope, brightness mean, roughness slope, and spectral centroid mean. B) Parameters for the vowel stimuli. There was a significant effect of emotion for pitch mean, spectral flux mean, brightness mean and slope, roughness slope, spectral centroid slope and inharmonicity mean. See main text for details. All units are arbitrary. Error bars correspond to the standard error of the mean. The slope sign refers to the general temporal ascending/descending tendency of the acoustical measure in stimuli.

Table 2
Results of the Bayesian mixed repeated measures ANOVAs on sentence stimuli, for correct categorization scores and intensity ratings. P(M): prior probability assigned to the model, P(M|data): probability of the model knowing the data, BF_M: Bayesian Factor of the model, BF_{10}: Bayesian Factor of the model compared to the null model.

	Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Correct categorization	Emotion	0.2	0.525	4.416	5.39	1.345
	Emotion + Group	0.2	0.274	1.509	2.813	1.403
	Null model (incl. subject)	0.2	0.097	0.431	1	
	Emotion + Group + Emotion × Group	0.2	0.054	0.228	0.553	1.004
	Group	0.2	0.05	0.211	0.515	1.261
Intensity	Emotion	0.2	0.679	8.479	3716.624	0.609
	Emotion + Group	0.2	0.278	1.542	1521.765	0.803
	Emotion + Group + Emotion x Group	0.2	0.042	0.176	230.261	2.025
	Null model (incl. subject)	0.2	1.828e −4	7.314e −4	1	
	Group	0.2	6.943e −5	2.777e −4	0.38	0.632

comparisons.

In a second step, based on the percentage of correct responses of the emotion categorization task for the vowel material, we performed a multifactorial analysis (MFA) using the package FactoMineR (Lê et al., 2008) in R to analyze the underlying data structure (emotional space). We could not perform a MFA for sentence material due to high performance (ceiling effect). For each participant (18 amusics and 18 controls), we built a matrix with the responses for each stimulus in each of the five-emotion categories (20 in total), equivalent to a contingency table. Each participant and each stimulus were projected in the multi-dimensional space recovered by the analysis. R2 (correlation factor) is retrieved for each participant for each dimension and then compared between groups with t-tests. The acoustic parameters describing the stimuli were correlated with the coordinates of the stimuli in the multidimensional space (r coefficients are obtained for the projection of each acoustic parameter in each dimension) to uncover the relation between emotion dimensions and acoustic parameters. The correlations of acoustic parameters with the dimensions were only descriptive and aimed to rank the degree of correlation of each variable for each dimension (Husson, 2013; Husson et al., 2008; Lê et al., 2008).

Table 3
Percent of answer types for each intended emotion for sentence material, averaged over the 18 participants for each group. Correct answers are on the diagonal.

	Answered / Expected	Joy	Sadness	Anger	Fear	Neutral
Amusics	Joy	**90.28**	2.78	0	6.94	0
	Sadness	0	**93.06**	0	6.94	0
	Anger	0	0	**100**	0	0
	Fear	0	6.94	8.33	**81.94**	2.78
	Neutral	0	8.33	1.39	0	**90.28**
Controls	Joy	**98.61**	0	0	1.39	0
	Sadness	1.39	**93.06**	0	5.56	0
	Anger	0	0	**98.61**	1.39	0
	Fear	0	2.78	4.17	**93.06**	0
	Neutral	0	2.78	1.39	0	**95.83**

3. Results

3.1. Sentences

Emotion categorization (Fig. 2). After comparison to the null model, the only model showing positive evidence was the one with only the main effect of emotion ($BF_{10} = 5.39$). This model was 1.91 times more likely than the model with the two main effects of group and emotion ($BF_{10} = 2.81$), which reached only anecdotic evidence compared to the null model. All other models showed no significant evidence ($BF_{10} < 1$). This was confirmed by a positive specific effect of emotion only ($BF_{inclusion} = 3.85$), and all other specific effects were not significant ($BF_{inclusion} < 0.5$) (Table 2). According to t-tests with Bonferroni correction, Anger was significantly better recognized than Fear ($t(34) = 3.55$, $pcorr<0.001$).

Within each participant group (amusic or control), no significant Pearson-correlation was found between the categorization performance and the MBEA score ($r(16) = 0.004$, $p = 0.99$ for controls, and $r(16) = 0.29$, $p = 0.24$ for amusics). Similarly, no correlation was found between categorization performance and the PDT ($r(16) = -0.26$, $p = 0.29$ for controls, and $r(16) = -0.43$, $p = 0.07$ for amusics). However, a correlation was found between categorization performance and PDT over the two participant groups ($r(34) = -0.43$, $p = 0.009$).

Confusion matrices are presented in Table 3 for the two participant groups. These matrices showed similar patterns for amusic and control participants.

Table 4
Results of the Bayesian mixed repeated measures ANOVAs on vowel stimuli, for correct categorization scores and intensity ratings. P(M): prior probability assigned to the model, P(M|data): probability of the model knowing the data, BF_M: Bayesian Factor of the model, BF_{10}: Bayesian Factor of the model compared to the null model.

	Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Correct categorization	Emotion + Group + Emotion x Group	0.2	0.682	8.583	211,784.75	1.288
	Emotion + Group	0.2	0.28	1.555	86,921.54	3.184
	Emotion	0.2	0.038	0.158	11,773.31	0.416
	Group	0.2	1.56E-05	6.22E-05	4.829	1.344
	Null model (incl. subject)	0.2	3.22E-06	1.29E-05	1	
Intensity	Emotion	0.2	0.571	5.327	1.09E+23	0.597
	Emotion + Group	0.2	0.38	2.448	7.21E+22	1.946
	Emotion + Group + Emotion x Group	0.2	0.049	0.207	9.35E+21	2.26
	Null model (incl. subject)	0.2	5.27E-24	2.11E-23	1	
	Group	0.2	2.24E-24	8.95E-24	0.425	1.191

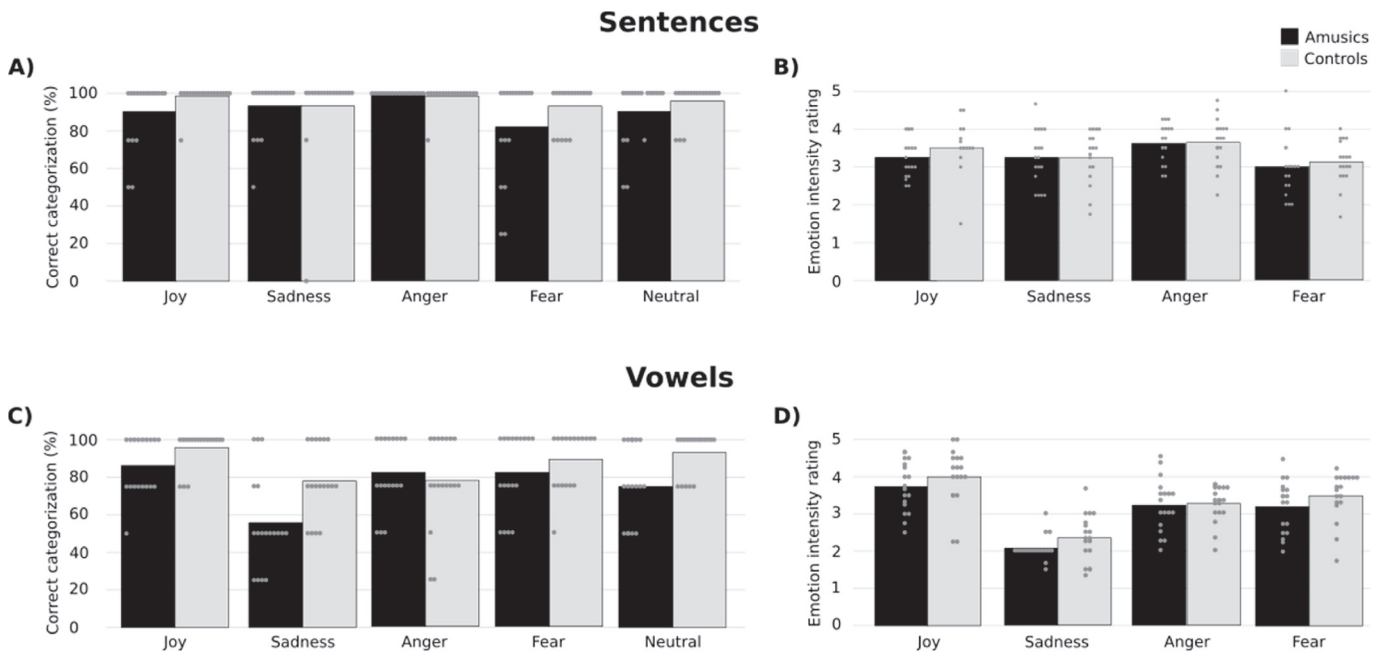

Fig. 2. Percentage of correct categorization for amusic and control groups for sentence (A) and vowel (C) material. Mean intensity ratings for amusic and control groups, for sentence (B) and vowel (D) material. Bars represent the group means and dots correspond to individual data points. Amusics showed normal emotion recognition for sentences compared to controls. A ceiling effect was observed for sentences material as a majority of the participants were at 100%. Amusics showed a deficit for emotion recognition in vowels compared to controls. Some amusics were better than others (100% of correct answers) and performance were correlated with the PDT in amusic participants, with lower scores for participant with large PDTs. There were no significant group differences in emotion intensity ratings.

Intensity ratings. The entire range (from 1 to 5) of intensity ratings was covered by the participants, showing that they fully used the subjective scale when rating the stimuli.

After comparison to the null model, the best model showing a decisive evidence was the one with the main effect of emotion only ($BF_{10} = 3716.62$) (Fig. 3). This model was 2.44 times more likely than the model with the two main effects of group and emotion ($BF_{10} = 1521.77$) and 16.16 times more likely than the model with the two main effects of emotion and group and their interaction ($BF_{10} = 230.26$). The model with the main effect of group showed no evidence compared to the null model ($BF_{10} = 0.38$). This was confirmed by a decisive specific effect of emotion only ($BF_{inclusion} = 2642.28$), and all other specific effects were non-significant ($BF_{inclusion} < 0.4$) (Table 2). According to t-tests with Bonferroni correction, Fear was rated significantly lower than Joy and Anger ($t(34) = 3.41$ pcorr = 0.012 and $t(34) = 5.58$ pcorr<0.001 respectively), and Sadness rated lower than Anger ($t(35) = 2.88$ pcorr = 0.042)[3]

Amusic and control participants rated the stimuli similarly: amusics' and controls' intensity ratings correlated across stimuli ($r(14) = 0.83$, $p < 0.001$)[4] Moreover, both groups had similar reliability for intensity judgments (α Cronbach amusics = 0.79 and α Cronbach controls = 0.88, Fisher-Bonett $Z = -1.08$, $p = 0.14$).

3.2. Vowels

Emotion categorization (Fig. 2). After comparison to the null model, the best model showing a decisive evidence was the one with the two main effects of group and emotion and their interaction ($BF_{10} = 211784.75$). This model was 2.44 times more likely than the model with the two main effects of group and emotion ($BF_{10} = 86921.54$) and 17.99 times more likely than the model with the main effect of emotion ($BF_{10} = 11773.31$). The model with the main effect of group showed positive evidence ($BF_{10} = 4.83$). This was confirmed by a decisive specific effect of emotion ($BF_{inclusion} = 35508.55$), a strong specific effect of group ($BF_{inclusion} = 16.91$) and positive evidence of the interaction between emotion and group ($BF_{inclusion} = 8.58$) (Table 4). According to t-tests with Bonferroni correction, Sadness was significantly less recognized than Joy, Fear, Anger and Neutrality (all pcorr<0.001). Amusics had lower recognition scores compared to controls only for Sadness ($t(16) = 2.76$ pcorr = 0.05) and for Neutrality ($t(16) = 3.42$ pcorr = 0.015) (other pcorr > 0.15).

Within each participant group (amusic or control), no significant Pearson-correlation was found between categorization performance and the MBEA score ($r(16) = 0.18$, $p = 0.49$ for controls, and $r(16) = 0.26$, $p = 0.29$ for amusics). Similarly, no correlation was found between categorization performance and the PDT ($r(16) = -0.070$, $p = 0.78$ for controls, and $r(16) = -0.32$, $p = 0.2$ for amusics). However, a correlation was found between categorization performance and the MBEA score over the two groups ($r(34) = 0.55$, $p < 0.001$). Similarly, a correlation was found between categorization performance and the PDT over the two groups ($r(34) = -0.50$, $p = 0.0021$).

For each participant group, response confusions are shown in confusion matrices in Table 5. Interestingly, amusics seemed to show more confusion of sadness with other emotions, in particular with neutrality, than did controls. Amusics seemed to have a bias in the recognition of sadness in vowels towards neutrality. This point will be further explored in the MFA reported below.

Intensity ratings. The entire range (from 1 to 5) of intensity ratings was covered by the participants, showing that they fully used the subjective scale to rate the stimuli.

After comparison to the null model, the best model showing decisive evidence was the one with the main effect of emotion only ($BF_{10} = 1.09e+23$) (Fig. 4). This model was 1.51 times more likely than the model with the two main effects of group and emotion ($BF_{10} = 7.22e+22$) and 11.6 times more likely than the model with the two main effects of emotion and group and their interaction ($BF_{10} = 9.35e+21$). The model with the main effect of group showed no evidence compared to the null model ($BF_{10} = 0.43$). This was confirmed by a decisive specific effect of emotion only ($BF_{inclusion} = 6e+15$), and all other specific effects were non-significant ($BF_{inclusion} < 0.5$) (Table 4). According to t-tests with Bonferroni correction, Sadness was rated significantly lower than Joy, Fear and Anger (all pcorr<0.001), and Anger and Fear rated lower than Joy (both pcorr<0.001)[5]

Amusic and control participants rated the stimuli similarly (correlation between groups across stimuli: $r(14) = 0.91$, $p < 0.001$)[6] Moreover, both groups had similar reliability for intensity judgements (α Cronbach amusics = 0.7 and α Cronbach controls = 0.76, Fisher-Bonett $Z = -0.43$, $p = 0.33$).

Multifactorial analysis. The multifactorial analysis revealed that four dimensions explained 69% percent of the variance in categorization data for vowels (Figure A1). The stimulus positions across these four dimensions confirmed the categorical structure of our stimulus set (Fig. 5A). In particular, the first dimension corresponded to the separation of anger and joy stimuli, the second dimension to the separation of anger and neutral/sadness/fear stimuli, the third dimension to the separation of fear stimuli compared to the other emotions, and the fourth dimension corresponded to the separation of neutral and sadness stimuli. We correlated the acoustic parameters with the emotional space dimensions resulting from the multifactorial analysis (Fig. 5B). Note in particular that the brightness mean ($r(18) = -0.6$) and slope ($r(18) = -0.49$), the spectral centroid mean ($r(18) = -0.56$) and slope ($r(18) = -0.46$) and the inharmonicity slope ($r(18) = -0.46$) of the stimuli correlated the most with the first dimension. The roughness mean, ($r(18) = 0.56$), the inharmonicity mean ($r(18) = 0.84$) and slope, ($r(18) = 0.48$), the spectral flux mean ($r(18) = 0.8$), and the spectral spread mean ($r(18) = -0.58$ and slope ($r(18) = -0.47$) of the stimuli correlated the most with the second dimension. The pitch mean ($r(18) = 0.82$) of the stimuli correlated the most with the third dimension. Both the brightness slope ($r(18) = 0.49$) and the spectral centroid slope

[3] We run another Bayesian ANOVA with intensity ratings of all the trials (correct or not). This confirmed the decisive evidence of the best model with the main effect of emotion (BF10 = 2.68.106). This model was 2.25 times more likely than the model with the two main effects of group and emotion (BF10 = 1.19.106) and 12.18 times more likely than the model with the two main effects of emotion and group and their interaction (BF10 = 223230). This was confirmed by a decisive specific effect of emotion only (BFinclusion = 1.95e+6), all other specific effects were non-significant (BFinclusion < 0.35). T tests with Bonferroni correction confirmed that Anger was rated higher than Sadness, Fear and Joy (all pcorr<0.03), and Joy rated higher than Fear (pcorr<0.001).

[4] We run another correlation measure with intensity ratings for all the trials (correct or not). It confirms that amusic and control participants rated the stimuli similarly: amusics' and control's intensity ratings correlated across stimuli ($r(14) = 0.84$, $p < 0.001$).

[5] We run another Bayesian ANOVA with intensity ratings of all the trials (correct or not). This showed a decisive evidence for the best model with the two main effects of emotion and group (BF10 = 1.56.1030), with amusics' ratings lower than the controls'. This model was 1.25 times more likely than the model with the main effect of emotion (BF10 = 1.25.1030) and 3.25 times more likely than the model with the two main effects emotion and group and their interaction (BF10 = 4.83.1029). This was confirmed by a decisive specific effect of emotion only (BFinclusion = ∞), all other specific effects were non-significant (BFinclusion < 1.09). T tests with Bonferroni correction confirmed that Joy was rated higher than Fear, Anger and Sadness (all pcorr<0.001), Fear and Anger rated higher than Sadness (all pcorr<0.001), Fear rated higher than Anger (pcorr<0.001).

[6] We run another correlation measure with intensity ratings for all the trials (correct or not). It confirms that amusic and control participants rated the stimuli similarly: amusics' and control's intensity ratings correlated across stimuli ($r(14) = 0.92$, $p < 0.001$).

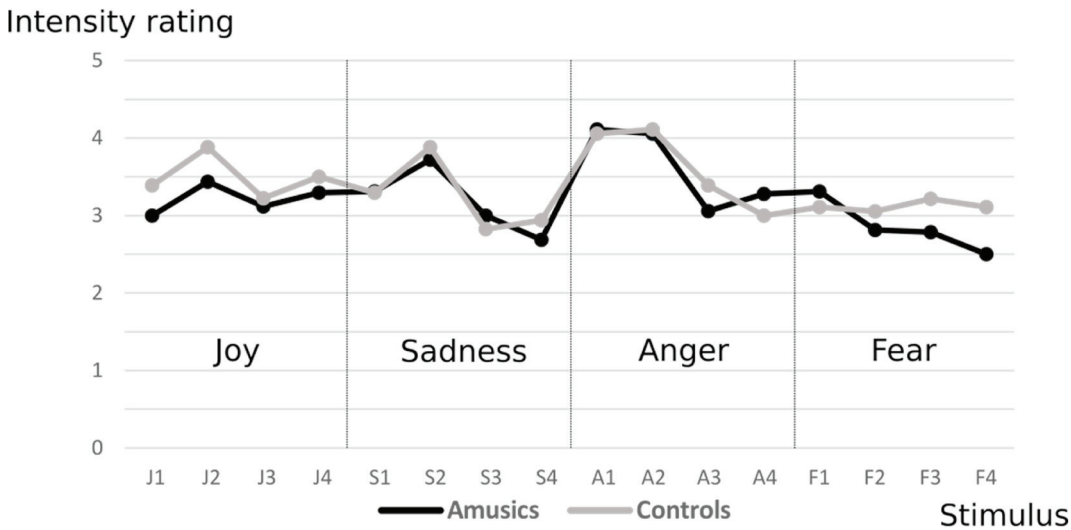

Fig. 3. Means of emotion intensity ratings by stimulus for the amusic and the control group, for sentence material. Stimuli are classified by emotion. Ratings are given on a scale from 1 (not intense) to 5 (very intense). J1-4: Joy, S1-4: Sadness, A1-4: Anger, F1-4: Fear.

Table 5
Percent of answer types for each intended emotion for vowel material, averaged over the 18 participants for each group. Correct answers are on the diagonal.

	Answered / Expected	Joy	Sadness	Anger	Fear	Neutral
Amusics	Joy	**86.11**	8.33	2.78	0	2.78
	Sadness	5.56	**55.56**	0	6.94	31.94
	Anger	2.78	0	**81.94**	6.94	8.33
	Fear	6.94	6.94	2.78	**81.94**	1.39
	Neutral	4.17	6.94	4.17	9.72	**75**
Controls	Joy	**95.83**	1.39	1.39	0	1.39
	Sadness	1.39	**77.78**	1.39	1.39	18.06
	Anger	1.39	5.56	**77.78**	8.33	6.94
	Fear	5.56	5.56	0	**88.89**	0
	Neutral	1.39	2.78	1.39	1.39	**93.06**

($r(18) = 0.56$) of the stimuli correlated the most with the fourth dimension. This suggests that numerous acoustic cues are used to distinguish anger from other emotions in vowels: the major acoustic cue used to distinguish fear from other emotions was pitch, whereas the major ones used to distinguish between neutral and sadness vowels were spectro-temporal cues, such as the slopes of brightness and spectral centroid. Interestingly, the third and fourth dimensions were the only dimensions to separate the two participant groups in terms of emotion classification (for dimension 3: $\cos2 = 0.259$ and $R2 = 0.81$ (±0.043) for controls, $\cos2 = 0.231$ and $R2 = 0.67$ (±0.06) for amusics, and for dimension 4: $\cos2 = 0.150$ and $R2 = 0.62$ (±0.054) for controls, $\cos2 = 0.068$ and $R2 = 0.36$ (±0.57) for amusics). The group comparison of individual R2 (Fig. 5C) on the third and the fourth dimensions revealed a significant between-group difference (dimension 3: $t(16) = 2.09$, $p = 0.044$, and dimension 4: $t(16) = 2.87$, $p = 0.007$), with control participants contributing more strongly to these dimensions.

3.3. Sentences and vowels

Emotion categorization. After comparison to the null model, the best model showing a decisive evidence was the one with the three main effects of group, emotion, and material (vowels or sentences) and the interaction between material and emotion (BF_{10} = 1.02e+13). This model was 1.89 times more likely than the model with the three main

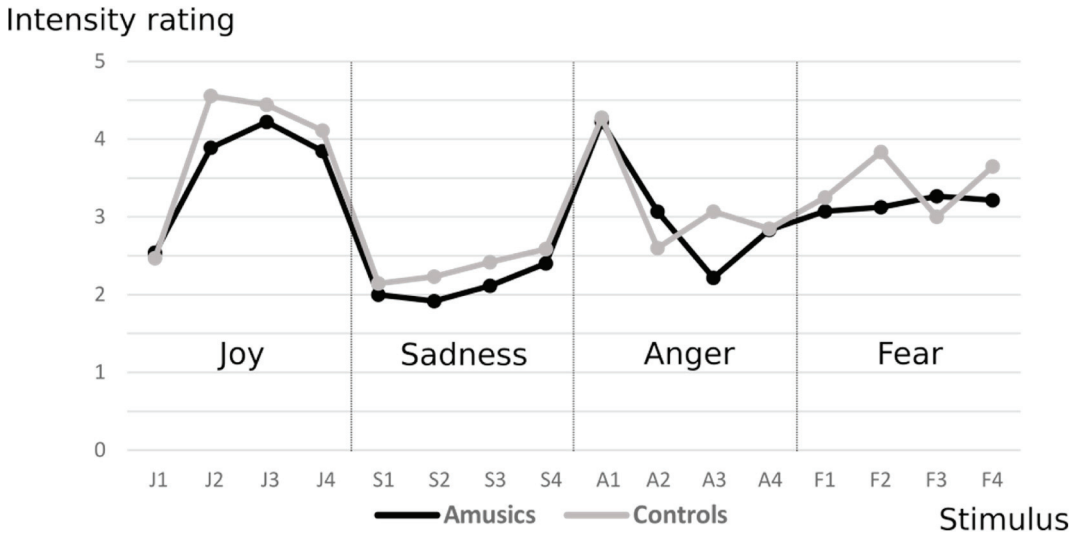

Fig. 4. Means of emotion intensity ratings by stimulus for the amusic and the control groups, for vowel material. Stimuli are classified by emotion. Ratings are given on a scale from 1 (not intense) to 5 (very intense). J1-4: Joy, S1-4: Sadness, A1-4: Anger, F1-4: Fear.

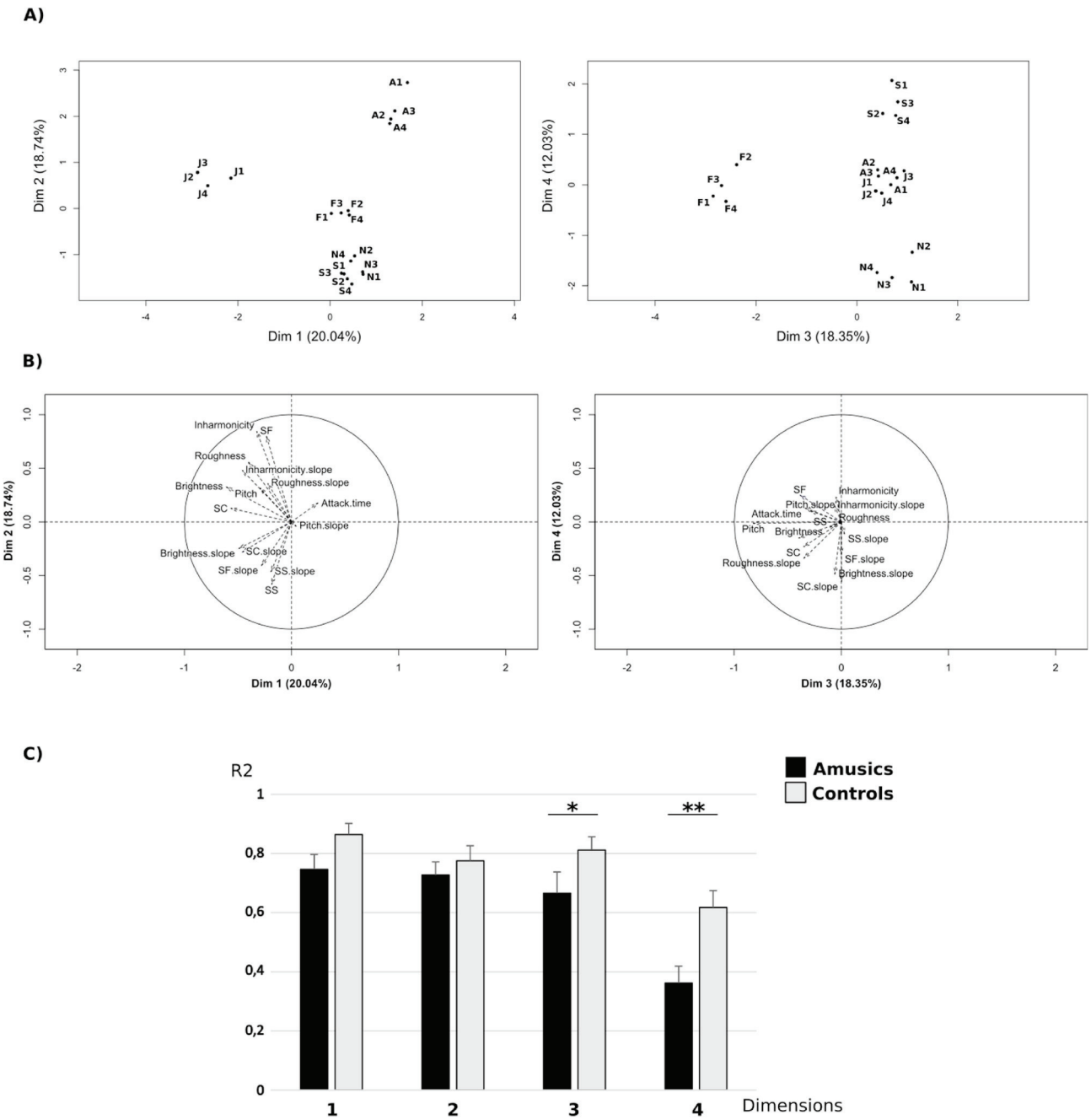

Fig. 5. Results of the multifactorial analysis (MFA) on categorization data for vowel stimuli. Categorization data were explained by a four-dimensional model. The third and fourth dimensions differentiate between the two groups (amusics and controls). The third dimension corresponds to the separation of fear stimuli from other emotions. The fourth dimension corresponds to the separation of neutral and sadness stimuli. A) Representation of the stimuli in the four dimensions. B) Mean R2 correlation of the two groups (amusics and controls) for the first four dimensions of the MFA. The groups were mainly separated by the fourth dimension ($p = 0.007$), and to a lesser extent by the third dimension ($p = 0.044$), error bars indicate standard error. C) Correlations between acoustic parameters describing the emotional stimuli with the four dimensions of the AFM. SF = Spectral Flux, SC = Spectral Centroid, SS = Spectral Spread. J1-4 = Joy, N1-4 = Neutral, S1-4 = Sadness, A1-4 = Anger, F1-4 = Fear.

effects of group, emotion and material and the interactions between material and group, and between material and emotion (BF_{10} = 5.41e+12). The best model was also 2.03 times more likely than the model with the three main effects of group, emotion and material and the interactions between material and emotion, and between group and emotion (BF_{10} = 5.02e+12). The best model was 3.25 times more likely than the model with the three main effects of group, emotion and material and the interactions between material and group, between material and emotion, and between group and emotion (BF_{10} = 3.14e+12). The best model was 4.81 times more likely than the model with the three main effects of group, emotion and material and the interactions between material and group, between material and emotion, between group and emotion and between material, group and emotion (BF_{10} = 2.12e+12). Finally, the best model was 5.34 times more likely than the model with the two main effects of emotion and material and the interaction between the two (BF_{10} = 1.91e+12). All other models were more than 2513 times less likely than the best model (all BF_{10}<4e+9). This was confirmed by a decisive evidence for the specific effects of material ($BF_{inclusion}$ = 7.44e+10), emotion ($BF_{inclusion}$ = 97881.46), and their interaction ($BF_{inclusion}$ = 5818.48), and a positive evidence for the

specific effect of group ($BF_{inclusion} = 4.83$). See supplementary data for the full result table (Table A1). Amusics had lower scores than Controls and Vowels were less well recognized than Sentences. According to t-tests with Bonferroni correction, Sadness was significantly less well recognized than Joy and Anger ($t(71) = 4.06$ pcorr<0.001 and $t(71) = 3.05$ pcorr = 0.03 respectively). Sadness ($t(35) = 4.76$ pcorr<0.001) and Anger ($t(35) = 5.39$ pcorr<0.001) were less well recognized in Vowels than in Sentences.

Intensity ratings. After comparison to the null model, the best model showing a decisive evidence was the one with the two main effects of emotion and material and the interaction between the two (BF_{10} = 3.92e+29). This model was 2.05 times more likely than the model with the three main effects of group, emotion and material and the interaction between material and emotion (BF_{10} = 1.91e+29). All other models were more than 10 times less likely than the best model (all BF_{10}<4e+28). This was confirmed by decisive evidence for the specific effects of material ($BF_{inclusion} = \infty$), emotion ($BF_{inclusion} = \infty$), and their interaction ($BF_{inclusion}$ = 9.76e+15), and no evidence for other main effects or interaction (all $BF_{inclusion}$ <0.22). See supplementary data for full table (Table A1). According to t-tests with Bonferroni correction, Joy was rated with stronger intensity than Fear and Sadness ($t(71) = 5.91$ pcorr<0.001 and $t(71) = 7.46$ pcorr<0.001 respectively), Anger rated as more intense than Sadness ($t(71) = 7.65$ pcorr<0.001), and Fear rated higher than Sadness ($t(71) = 4.21$ pcorr<0.001). Joy was rated as more intense in vowels than sentences ($t(35) = 3.78$ pcorr<0.001), and Sadness and Anger rated higher in sentences than vowels ($t(35) = 9.42$ pcorr<0.001 and $t(35) = 3.37$ pcorr = 0.01 respectively).

4. Discussion

The aim of the present study was to investigate emotional prosody perception in congenital amusia. To do so, we used sentences with neutral semantic content and short vowels (/a/) to investigate the explicit recognition of emotions and related intensity ratings. The stimuli were pronounced to express four different basic emotions: joy, fear, anger, and sadness, and with a neutral voice. To further investigate the perception of emotions in speech, and based on recent findings about congenital amusia and music-evoked emotions (Lévêque et al., 2018), participants were also asked to rate the intensity of the emotion previously recognized. Overall, the results revealed an explicit deficit in the amusic group for emotion recognition in vowels, but not in sentences. This deficit was most pronounced for sadness and neutrality, which tended to be confused with each other or with fear more often by amusics than by controls for the vowel material. These impairments in emotional prosody recognition seem to be linked to a deficit in the processing of pitch and spectro-temporal parameters of the stimuli, with a relative preservation of harmonicity and roughness processing in congenital amusia. Despite this explicit deficit, implicit judgments on these emotions seemed to be preserved, as intensity ratings were similar in both participant groups for both materials.

4.1. Deficit of emotional prosody processing in congenital amusia

For the sentence material, amusics' performance did not differ from controls' performance: both groups were equally able to recognize the expressed emotions. This was confirmed by the similar confusion matrices between the two participant groups. According to our acoustic analysis on the sentence material, we found significant differences between emotions for a variety of pitch-related cues, spectral cues and spectro-temporal cues (slopes of spectral parameters variation across time). We suggest that these numerous acoustic cues associated with the rather long duration of our stimuli (an average duration of 1470 ms per sentence) were sufficient for participants of the two groups to distinguish emotions. The combination (and accumulation over time) of these acoustic cues could explain why amusics' emotion recognition was similar to that of controls, as previously shown by Lolli et al. (2015). However, in the study of Thompson et al. (2012), amusic participants showed difficulties to detect emotions in spoken sentences, in particular for happy, tender, irritated and sad stimuli. This difference in result to our findings on sentence material could be due to several reasons: (1) the difference of recruitment of the amusic participants in the two studies; in particular, the entire MBEA test was used to detect amusia in our study, whereas Thompson et al. (2012) only used the three pitch subtests of the MBEA. As shown in their results, the categorization scores were correlated to the vocal pitch contour of their stimuli. As amusics were selected in their study only on their deficit of pitch, it could explain why they performed poorly in their sentence material. In comparison, in the present study, selection of the amusics was less not focused on the pitch dimension only (our criteria also included possible deficits in the other three subtests of the MBEA, notably testing for meter, rhythm and memory). This could contribute to explain why some of the amusics performed better with the sentences material. 2) It might also be possible that emotional prosody information in their material was less supported by other acoustical features (other than pitch), and/or that the differences in pitch were not as pronounced as in our material leading to more perceptual difficulties to detect an emotion. 3) An additional explanation could be that Thompson et al. used a paradigm with six emotions, leading to possibly more confusions between the emotion categories than our task (with four emotions and neutrality). Overall, it appears that results obtained with sentence materials across different previous studies (including ours, Thompson et al., 2012 as well as Lolli et al. (2015)) were not able to clearly pinpoint a deficit in amusia, highlighting the need to use also shorter, simpler verbal material to investigate emotional processing.

The results for the full-sentence stimuli showed no evidence for an impairment in emotion recognition in amusia. Note that near-ceiling performance was observed in both participant groups. To further investigate possible group differences, we also presented participants with a simpler 400 ms vowel material where temporal cues are minimal, and less acoustic information is available than in full sentences. For the vowel material, we observed a group effect, with lower performance in the amusics relative to the controls. Interestingly, this was accompanied by an interaction between group and emotion. Specifically, amusics had more difficulties recognizing sadness and neutrality than did controls. These results were confirmed by the confusion matrices of each group, showing more confusion between sadness and neutrality in the amusic group than in the control group (these two emotions being also sometimes misclassified as fear). In controls, this bias was less marked, and controls had more difficulties recognizing anger than sadness, which was not the case for amusics. A multifactorial analysis confirmed that all stimuli were well separated across the five emotions. This analysis also revealed how acoustic parameters might influence emotion categorization. Correlations between acoustic parameters and the fourth axis of the emotional space (recovered by the multifactorial analysis) separating sadness and neutral stimuli revealed the similarity of several acoustic features between neutral and sadness stimuli, the most frequent confusion in the amusic group. The largest correlations, meaning that the acoustic feature could theoretically be used to distinguish the two emotions, occurred for (1) brightness slope and (2) spectral centroid slope. These spectro-temporal cues have previously been shown to be perceptually less salient than other acoustic cues in non-amusic, normal-hearing listeners (Caclin et al., 2005), showing that neutral and sadness stimuli are perceptually rather close. This suggests that amusic individuals have difficulty detecting contrasts between emotions with less salient variations of acoustic parameters. Indeed, the differences between neutrality and sadness are minimal in pitch, brightness, roughness, and inharmonicity, which are often used by non-amusic participants to recognize an emotion (Marin et al., 2015).

The first two dimensions of the multifactorial analysis correlated with several acoustic cues, such as roughness and inharmonicity. No deficit was observed in the amusic group on these two dimensions. This confirmed previous results showing that amusic individuals judge

affective sounds mostly based on roughness and sometimes on harmonicity, whereas, relative to controls, they have reduced sensitivity to other acoustic parameters (Marin et al., 2015). The third dimension of the multifactorial analysis correlated the highest with pitch, and the results of the amusic and control groups differ significantly for this dimension, confirming the amusic deficit in emotional prosody recognition when pitch is needed (Liu et al., 2015; Marin et al., 2015). The findings suggest that amusics need more variation of acoustic cues in speech to understand emotional prosody and rely more on accumulation of acoustic cues over time to determine the emotion.

Our additional analysis for sentence and vowel materials revealed a significant interaction between emotion and material showing that, depending on the material (sentences or vowels), participants did not recognize all emotions in the same way. Post hoc tests showed that sadness was less well recognized in vowels than in sentences, possibly revealing the need to process subtler acoustic cues and the less dynamic cues in sad vowel stimuli compared to sentences. The best model of the analysis showed that amusic participants had lower recognition scores overall. However, the second best model of our Bayesian analysis included the interaction of group and material, showing a larger deficit of amusics with vowel material relative to other group/material combinations. Even if this interaction was not present in the best model, the second model included this interaction and it was only 1.89 times less likely than the best model. Overall, there might thus be a general deficit of amusic individuals to recognize emotional prosody, and this deficit might be less detectable in sentences as amusics can compensate with other mechanisms or benefit from more acoustic information available over time to detect the emotion. Hence, for vowel processing with less acoustic information available, this deficit might be too strong for amusics to be compensated and is revealed by our data. These findings suggest a gradual difficulty of detecting emotional prosody in relation with the length and the acoustic cues available in the to-be-processed material. The results of the present study reveal the need for further studies on emotional prosody in amusia, which should systematically manipulate the length and the amount of available acoustic features of the stimuli to further investigate this hypothesis.

4.2. Preserved implicit processing of emotional prosody in congenital amusia

To determine whether congenital amusic individuals were also impaired for implicit processing of emotional prosody, participants were asked to provide intensity ratings to each perceived emotion. Previous studies with patients with acquired amusia revealed a possible dissociation of emotion processing between recognition and the intensity ratings (Hirel et al., 2014; Peretz et al., 1998). These studies showed a spared recognition of musical emotions whereas intensity ratings were impaired. Interestingly, here we observed unimpaired performance for the intensity ratings in congenital amusia, both for sentence and vowel stimuli. Indeed, we observed similar intensity ratings in both participant groups, with ratings differing only between emotions. Furthermore, the analysis on intensity ratings with sentences and vowels together revealed only two main effects of emotion and material, along with no group effect or interaction. The correlation of intensity ratings across the stimuli revealed that both groups rated the stimuli in a similar way across emotions. This correlation reveals that amusic participants' intensity ratings of the stimuli were similar to those of the controls, regardless of the material used. Even if the task might have involved some memory load for the participants (they had to retain the stimulus to give their intensity rating after their recognition response), and some previous research has suggested that amusics have a memory deficit specific for pitch (Albouy et al., 2016, 2015; Albouy et al., 2013a, 2013b; Tillmann et al., 2016b; Tillmann et al., 2009), amusics' participants demonstrated no deficit for intensity ratings. Furthermore, the results showed clear differences between intensity ratings for the different emotions. These differences between emotions were more pronounced in vowels than in sentences. These observations suggest that amusic participants were able to rate the intensity of the different emotions, even if they did not recognize them as well as did the controls. Results obtained with intensity judgements on all stimuli, including those not correctly categorized, showed the same pattern of results with a main effect of emotion. A small effect of group was revealed in the vowel intensity analysis, but the $BF_{inclusion}$ of the group factor was very small (<3, which can be referred to as anecdotal, following the labeling suggested by Wagenmakers et al. (2018)), revealing a non-significant effect of group in the model and a stronger effect of emotion. In this complementary analysis on intensity ratings on all stimuli, confusion between sad and neutral stimuli by amusic participants could have biased the result pattern and could explain why amusics' intensity ratings were lower than those of controls. For this reason, the analysis focusing on correct trials only is more accurate to point out the specific implicit processing of emotional prosody in both participant groups.

These analyses confirmed that emotion categorization and intensity ratings can be at least partially separated as two processes within global emotion processing as has been previously described in patients with acquired amusia or musical anhedonia (Hirel et al., 2014; Peretz et al., 1998), and that the deficit in amusia occurs mainly in explicit emotion categorization, whereas implicit emotional prosody processing is relatively preserved. Moreover, a previous study with congenital amusics revealed preserved intensity ratings of musical emotions whereas the categorization of these emotions was disturbed (Lévêque et al., 2018). As described in previous research on amusia, implicit and explicit investigation methods might be measuring the same internal representation of a stimulus. However, implicit methods would rather need lower activation levels of this potential representation, with less precise knowledge about this stimulus, compared to explicit methods. Intensity ratings of an emotional stimulus might also need rather weaker representations of the stimulus whereas the categorization would need higher activation levels and conscious representation of the stimulus. This difference between implicit and explicit processing and mental representation is referred to by Cleeremans and Jimenez (2002) as "a difference in degree rather than in kind" (Cleeremans and Jimenez, 2002; Tillmann et al., 2012; Tillmann et al., 2007). Overall, these previous studies confirmed the potential use of intensity ratings as a reliable measure of implicit processing of emotions, suggesting that this rating can be done without having to consciously categorize the intended emotion. Previous results support this conclusion: in a recent study with amusic participants performing explicit tasks on musical material, the results suggested explicit tonal processing difficulties alongside preserved implicit processing (Tillmann et al., 2016a). Indeed, when explicitly asked to judge the tonality of musical material, amusic participants had difficulties, but when the judgement was implicit, they performed as well as did controls (Tillmann et al., 2016a; Zendel et al., 2015). Similarly, our findings on verbal material showed some impairment of explicit recognition and categorization of emotion in vowels, but preserved implicit knowledge for the processing of the intensity of the emotional content in amusia. These results are in line with previous reports showing that the perception of emotional prosody does not only involve a fronto-temporal network, which is impaired in congenital amusia, (Albouy et al., 2013a; Hyde et al., 2007, 2006, 2011), but also extend to other regions, such as probably the amygdala or the insula (Frühholz et al., 2016), which would be preserved in congenital amusia.

In sum, previous findings have suggested preserved implicit processing capacities in congenital amusia for musical materials (Lévêque et al., 2018; Marin et al., 2015; Moreau et al., 2009; Omigie et al., 2013; Peretz et al., 2009; Tillmann et al., 2014, 2012; Tillmann et al., 2016a), and the present study expands these findings to a new domain, suggesting preserved implicit processing of pitch and spectro-temporal auditory features for verbal materials as well. This preserved implicit processing could provide a basis on which to develop new methods for auditory rehabilitation in amusia.

4.3. Understanding the deficit of congenital amusia beyond impaired pitch processing in musical perception

Various acoustic cues differentiate the emotions of the stimuli in the present study. For the vowels, the variations in acoustic parameters could be linked with the differences in recognition accuracy between groups. In the multifactorial analysis, pitch was correlated with the third dimension, which was where the amusic group's performance differed significantly from that of the controls. This provided support for the hypothesis of a pitch perception deficit in amusia (Liu et al., 2015; Marin et al., 2015). In addition, the fourth dimension of the multifactorial analysis revealed a significant difference between amusics and controls and correlated with spectro-temporal variations of the stimuli, namely the slopes of the acoustic parameters, in particular brightness and spectral centroid. This observation suggests that amusics could have a subtle deficit in the perception of these acoustic parameters, leading to a deficit in perception of a particular set of aurally-presented emotions. This deficit in spectro-temporal variation processing in amusia leads us to re-evaluate the potential pitch-perception deficit (Zhang et al., 2017). Indeed, several studies have highlighted a contour-related pitch-perception deficit in amusia (Foxton et al., 2004; Patel et al., 2005). In agreement with these findings, the results of the multifactorial analysis seem to suggest a deficit in the processing of dynamic spectral information in congenital amusia. This is also in keeping with the deficit in short-term memory for pitch sequences, which characterizes amusic participants (Tillmann et al., 2016b; Whiteford and Oxenham, 2017). Finally, comparing results from perception of emotion in both sentences and vowels also confirmed that when speech prosody stimuli are long enough, accumulation of evidence allows the amusics to perform as well as controls for recognition of emotions (Lolli et al., 2015).

Moreover, our results lead to a better understanding of pitch processing in congenital amusia beyond musical material. Indeed, congenital amusia was first described as a music processing impairment (Ayotte et al., 2002; Peretz et al., 2003). However, recently, evidence has begun to accumulate that indicates that the pitch deficit in amusia extends to verbal materials as well (Nguyen et al., 2009; Tillmann et al., 2011a, 2011b; Zhang et al., 2017). This means that congenital amusia is a deficit not only of music perception but also of speech processing, in particular emotional and intentional prosody as well as tonal language, when pitch carries relevant information (Lolli et al., 2015; Nguyen et al., 2009; Patel et al., 2008; Thompson et al., 2012). Even though this deficit is small, here only observed with vowels, we also observed a difference in the experience of music and speech of amusics in their everyday life, compared to controls. Similar to what has been shown before, the musical emotion questions in the questionnaire revealed that amusic individuals less frequently experience emotions when listening to music (McDonald and Stewart, 2008; Omigie et al., 2012). More interestingly, the speech portion of the questionnaire showed a deficit in amusia for processing speech in a context of subtle acoustic variations (such as accent, intentional prosody). Even if this questionnaire was not specific for emotional prosody, it did reveal a small deficit for speech in amusia, beyond music, that could impact everyday life. For example, this impairment could become problematic when speech is presented in degraded conditions, such as speech in noise, speech in speech, or speech over the phone. These conditions of hearing have been shown to challenge comprehension in normal-hearing participants (Oxenham, 2008, 2012), particularly for the elderly, and could be even more difficult for amusic participants.

5. Conclusion

Overall, the results of the present study allow for a better understanding of the perceptual deficits of congenital amusia regarding speech material. They revealed a deficit in emotional prosody processing in amusia by examining emotion recognition of sentences and short vowels. The results suggest that the use of acoustic-constrained material can reveal subtle deficits in congenital amusia and allow for a better definition of acoustic cues that are critical for emotion perception in speech and its impairment in amusic individuals. The results also demonstrate a dissociation between explicit and implicit processing of emotions in congenital amusia (Lévêque et al., 2018). This finding contributes to a better understanding of the complex relationship between music and emotional prosody processing, and provides further elements for the comprehension of fine acoustic structures underlying music and speech appraisal.

Declaration of competing interest

None.

Acknowledgments

This work was supported by a grant from the Research National Agency (ANR-11-BSH2-001-01 to B. Tillmann and A. Caclin. This work was conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060) and of the LabEx Cortex ("Construction, Function and Cognitive Function and Rehabilitation of the Cortex", ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). This work was realized as part of a Marie Curie fellowship (International Incoming Fellowship) of A. Bhatara. We would like to thank Céline Tourlonnias for pretests with the sentence material. We also thank Lison Fanuel for her help with the Bayesian Analyses.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2019.107234.

References

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., Peretz, I., 2016. Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. Sci. Rep. 6, 18861. https://doi.org/10.1038/srep18861.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., et al., 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain: J. Neurol. 136 (Pt 5), 1639–1661. https://doi.org/10.1093/brain/awt082.

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., Caclin, A., 2015. Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data. Front. Hum. Neurosci. 9, 20. https://doi.org/10.3389/fnhum.2015.00020.

Albouy, P., Schulze, K., Caclin, A., Tillmann, B., 2013. Does tonality boost short-term memory in congenital amusia? Brain Res. 1537, 224–232. https://doi.org/10.1016/j.brainres.2013.09.003.

Allen, E.J., Moerel, M., Lage-Castellanos, A., De Martino, F., Formisano, E., Oxenham, A. J., 2018. Encoding of natural timbre dimensions in human auditory cortex. Neuroimage 166, 60–70. https://doi.org/10.1016/j.neuroimage.2017.10.050.

Ayotte, J., Peretz, I., Hyde, K., 2002. Congenital amusiaA group study of adults afflicted with a music-specific disorder. Brain 125 (2), 238–251. https://doi.org/10.1093/brain/awf028.

Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., Belin, P., 2014. Adaptation to vocal expressions reveals multistep perception of auditory emotion. J. Neurosci.: Off. J. Soc. Neurosci 34 (24), 8098–8105. https://doi.org/10.1523/JNEUROSCI.4820-13.2014.

Caclin, A., McAdams, S., Smith, B.K., Winsberg, S., 2005. Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J. Acoust. Soc. Am. 118 (1), 471–482.

Charpentier, J., Kovarski, K., Roux, S., Houy-Durand, E., Saby, A., Bonnet-Brilhault, F., et al., 2018. Brain mechanisms involved in angry prosody change detection in school-age children and adults, revealed by electrophysiology. Cognit. Affect Behav. Neurosci. https://doi.org/10.3758/s13415-018-0602-8.

Cleeremans, A., Jimenez, L., 2002. Implicit learning and consciousness: a graded, dynamic perspective. In: French, R.M., Cleeremans, A. (Eds.), Implicit Learning and Consciousness: an Empirical. Psychology Press.

Escoffier, N., Zhong, J., Schirmer, A., Qiu, A., 2013. Emotional expressions in voice and music: same code, same effect? Hum. Brain Mapp. 34 (8), 1796–1810. https://doi.org/10.1002/hbm.22029.

Foxton, J.M., Dean, J.L., Gee, R., Peretz, I., Griffiths, T.D., 2004. Characterization of deficits in pitch perception underlying 'tone deafness.'. Brain 127 (4), 801–810. https://doi.org/10.1093/brain/awh105.
Foxton, J.M., Nandy, R.K., Griffiths, T.D., 2006. Rhythm deficits in 'tone deafness'. Brain Cogn. 62 (1), 24–29. https://doi.org/10.1016/j.bandc.2006.03.005.
Frühholz, S., Ceravolo, L., Grandjean, D., 2012. Specific brain networks during explicit and implicit decoding of emotional prosody. Cerebr. Cortex 22 (5), 1107–1117. https://doi.org/10.1093/cercor/bhr184.
Frühholz, S., Trost, W., Kotz, S.A., 2016. The sound of emotions-Towards a unifying neural network perspective of affective sound processing. Neurosci. Biobehav. Rev. 68, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002.
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.
Gosselin, N., Paquette, S., Peretz, I., 2015. Sensitivity to musical emotions in congenital amusia. Cortex J. Devoted Stud. Nerv. Syst. Behav. 71, 171–182. https://doi.org/10.1016/j.cortex.2015.06.022.
Hirel, C., Lévêque, Y., Deiana, G., Richard, N., Cho, T.-H., Mechtouff, L., Nighoghossian, N., 2014. Acquired amusia and musical anhedonia. Rev. Neurol. 170 (8–9), 536–540. https://doi.org/10.1016/j.neurol.2014.03.015.
Husson, 2013. Tools to interpret the results obtained by principal component methods (PCA, MCA, MFA). Retrieved from. https://www.youtube.com/watch?v=Uhw-1NilmAk.
Husson, F., Josse, J., Lê, S., 2008. FactoMineR: exploratory multivariate data analysis with R. Retrieved January 28, 2019, from. http://factominer.free.fr/index.html.
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. J. Neurosci. 27 (47), 13028–13032. https://doi.org/10.1523/JNEUROSCI.3039-07.2007.
Hyde, K.L., Peretz, I., 2004. Brains that are out of tune but in time. Psychol. Sci. 15 (5), 356–360. https://doi.org/10.1111/j.0956-7976.2004.00683.x.
Hyde, K.L., Zatorre, R.J., Griffiths, T.D., Lerch, J.P., Peretz, I., 2006. Morphometry of the amusic brain: a two-site study. Brain: J. Neurol. 129 (Pt 10), 2562–2570. https://doi.org/10.1093/brain/awl204.
Hyde, K.L., Zatorre, R.J., Peretz, I., 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. Cerebr. Cortex 21 (2), 292–299. https://doi.org/10.1093/cercor/bhq094 (New York, N.Y.: *1991*).
Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2010. Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. Neuropsychologia 48 (9), 2630–2639. https://doi.org/10.1016/j.neuropsychologia.2010.05.009.
Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2012. Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Mem. Cogn. 40 (7), 1109–1121. https://doi.org/10.3758/s13421-012-0208-2.
Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., Gruber, O., 2009. Functional architecture of verbal and tonal working memory: an FMRI study. Hum. Brain Mapp. 30 (3), 859–873. https://doi.org/10.1002/hbm.20550.
Lartillot, O., Toiviainen, P., 2007. Mir in Matlab (ii): A Toolbox for Musical Feature Extraction from Audio. Austrian Computer Society.
Lê, S., Rennes, A., Josse, J., Rennes, A., Husson, F., Rennes, A., 2008. FactoMineR: an R package for multivariate analysis. J. Stat. Softw. 1–18.
Lee, M.D., Wagenmakers, E.-J., 2014. Bayesian cognitive modeling: a practical course. .
Leveque, Y., Fauvel, B., Groussard, M., Caclin, A., Albouy, P., Platel, H., Tillmann, B., 2016. Altered intrinsic connectivity of the auditory cortex in congenital amusia. J. Neurophysiol. https://doi.org/10.1152/jn.00663.2015 jn.00663.2015.
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. https://doi.org/10.1037/neu0000461.
Liebenthal, E., Silbersweig, D.A., Stern, E., 2016. The Language, tone and prosody of emotions: neural substrates and dynamics of spoken-word emotion perception. Front. Neurosci. 10 https://doi.org/10.3389/fnins.2016.00506.
Lima, C.F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S.K., Warren, J.D., Stewart, L., 2016. Impaired socio-emotional processing in a developmental music disorder. Sci. Rep. 6 https://doi.org/10.1038/srep34911.
Liu, F., Jiang, C., Francart, T., Chan, A.H.D., Wong, P.C.M., 2017. Perceptual learning of pitch direction in congenital amusia: evidence from Chinese speakers. Music Percept. 34 (3), 335–351.
Liu, F., Jiang, C., Wang, B., Xu, Y., Patel, A.D., 2015. A music perception disorder (congenital amusia) influences speech comprehension. Neuropsychologia 66, 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001.
Lolli, S.L., Lewenstein, A.D., Basurto, J., Winnik, S., Loui, P., 2015. Sound frequency affects speech emotion perception: results from congenital amusia. Front. Psychol. 6, 1340. https://doi.org/10.3389/fpsyg.2015.01340.
Loui, P., Alsop, D., Schlaug, G., 2009. Tone deafness: a new disconnection syndrome? J. Neurosci.: Off. J. Soc. Neurosci 29 (33), 10215–10220. https://doi.org/10.1523/JNEUROSCI.1701-09.2009.
Marin, M.M., Thompson, W.F., Gingras, B., Stewart, L., 2015. Affective evaluation of simultaneous tone combinations in congenital amusia. Neuropsychologia 78, 207–220. https://doi.org/10.1016/j.neuropsychologia.2015.10.004.
McDonald, C., Stewart, L., 2008. Uses and functions of music in congenital amusia. Music Percept. 25 (4), 345. https://doi.org/10.1525/MP.2008.25.4.345.
Moreau, P., Jolicoeur, P., Peretz, I., 2009. Automatic brain responses to pitch changes in congenital amusia. Ann. N. Y. Acad. Sci. 1169, 191–194. https://doi.org/10.1111/j.1749-6632.2009.04775.x.
Moreau, P., Jolicœur, P., Peretz, I., 2013. Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. Brain Cogn. 81 (3), 337–344. https://doi.org/10.1016/j.bandc.2013.01.004.
Nan, Y., Huang, W.T., Wang, W.J., Liu, C., Dong, Q., 2016. Subgroup differences in the lexical tone mismatch negativity (MMN) among Mandarin speakers with congenital amusia. Biol. Psychol. 113, 59–67. https://doi.org/10.1016/j.biopsycho.2015.11.010.
Nguyen, S., Tillmann, B., Gosselin, N., Peretz, I., 2009. Tonal language processing in congenital amusia. Ann. N. Y. Acad. Sci. 1169, 490–493. https://doi.org/10.1111/j.1749-6632.2009.04855.x.
Omigie, D., Müllensiefen, D., Stewart, L., 2012. The experience of music in congenital amusia. Music Percept. Interdiscip. J. 30 (1), 1–18. https://doi.org/10.1525/mp.2012.30.1.1.
Omigie, D., Pearce, M.T., Williamson, V.J., Stewart, L., 2013. Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia 51 (9), 1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010.
Oxenham, A.J., 2008. Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. Trends Amplif. 12 (4), 316–331. https://doi.org/10.1177/1084713808325881.
Oxenham, A.J., 2012. Pitch perception. J. Neurosci.: Off. J. Soc. Neurosci 32 (39), 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012.
Patel, A.D., Foxton, J.M., Griffiths, T.D., 2005. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. Brain Cogn. 59 (3), 310–313. https://doi.org/10.1016/j.bandc.2004.10.003.
Patel, A.D., Wong, M., Foxton, J., Lochy, A., Peretz, I., 2008. Speech intonation perception deficits in musical tone deafness (congenital amusia). Music Percept. Interdiscip. J. 25 (4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357.
Peretz, I., 2016. Neurobiology of congenital amusia. Trends Cogn. Sci. 20 (11), 857–867. https://doi.org/10.1016/j.tics.2016.09.002.
Peretz, I., Brattico, E., Järvenpää, M., Tervaniemi, M., 2009. The amusic brain: in tune, out of key, and unaware. Brain: J. Neurol. 132 (Pt 5), 1277–1286. https://doi.org/10.1093/brain/awp055.
Peretz, I., Champod, A.S., Hyde, K., 2003. Varieties of musical disorders. The montreal battery of evaluation of amusia. Ann. N. Y. Acad. Sci. 999, 58–75.
Peretz, I., Cummings, S., Dubé, M.-P., 2007. The genetics of congenital amusia (tone deafness): a family-aggregation study. Am. J. Hum. Genet. 81 (3), 582–588.
Peretz, I., Gagnon, L., Bouchard, B., 1998. Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition 68 (2), 111–141. https://doi.org/10.1016/S0010-0277(98)00043-2.
Peretz, I., Vuvan, D.T., 2017. Prevalence of congenital amusia. Eur. J. Hum. Genet.: EJHG (Eur. J. Hum. Genet.) 25 (5), 625–630. https://doi.org/10.1038/ejhg.2017.15.
Pfeuty, M., Peretz, I., 2010. Abnormal pitch-time interference in congenital amusia: evidence from an implicit test. Attention. Percept. Psychophys. 72 (3), 763–774. https://doi.org/10.3758/APP.72.3.763.
Sloboda, J.A., Wise, K.J., Peretz, I., 2005. Quantifying tone deafness in the general population. Ann. N. Y. Acad. Sci. 1060 (1), 255–261. https://doi.org/10.1196/annals.1360.018.
Stewart, L., 2011. Characterizing congenital amusia. Q. J. Exp. Psychol. 64 (4), 625–638. https://doi.org/10.1080/17470218.2011.552730, 2006.
Thompson, W.F., Marin, M.M., Stewart, L., 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. Proc. Natl. Acad. Sci. 109 (46), 19027–19032. https://doi.org/10.1073/pnas.1210344109.
Tillmann, B., Albouy, P., Caclin, A., 2015. Congenital amusias. Handb. Clin. Neurol. 129, 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0.
Tillmann, B., Albouy, P., Caclin, A., Bigand, E., 2014. Musical familiarity in congenital amusia: evidence from a gating paradigm. Cortex J. Devoted Stud. Nerv. Syst. Behav. 59, 84–94. https://doi.org/10.1016/j.cortex.2014.07.012.
Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., Peretz, I., 2011. Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. Front. Psychol. 2, 120. https://doi.org/10.3389/fpsyg.2011.00120.
Tillmann, B., Gosselin, N., Bigand, E., Peretz, I., 2012. Priming paradigm reveals harmonic structure processing in congenital amusia. Cortex J. Devoted Stud. Nerv. Syst. Behav. 48 (8), 1073–1078. https://doi.org/10.1016/j.cortex.2012.01.001.
Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., Bigand, E., 2016. Discrimination of tonal and atonal music in congenital amusia: the advantage of implicit tasks. Neuropsychologia 85, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027.
Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., Caclin, A., 2016. Impaired short-term memory for pitch in congenital amusia. Brain Res. 1640 (Part B), 251–263. https://doi.org/10.1016/j.brainres.2015.10.035.
Tillmann, B., Peretz, I., Bigand, E., Gosselin, N., 2007. Harmonic priming in an amusic patient: the power of implicit tasks. Cogn. Neuropsychol. 24 (6), 603–622. https://doi.org/10.1080/02643290701609527.
Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., Peretz, I., 2011. Fine-grained pitch processing of music and speech in congenital amusia. J. Acoust. Soc. Am. 130 (6), 4089–4096. https://doi.org/10.1121/1.3658447.
Tillmann, B., Schulze, K., Foxton, J.M., 2009. Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. Brain Cogn. 71 (3), 259–264. https://doi.org/10.1016/j.bandc.2009.08.003.
Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Morey, R.D., 2017. Bayesian inference for psychology. Part II: example applications with JASP. Psychon. Bull. Rev. 1–19 https://doi.org/10.3758/s13423-017-1323-7.
Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Morey, R.D., 2018. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon. Bull. Rev. 25 (1), 35–57. https://doi.org/10.3758/s13423-017-1343-3.

Whiteford, K.L., Oxenham, A.J., 2017. Auditory deficits in amusia extend beyond poor pitch perception. Neuropsychologia 99, 213–224. https://doi.org/10.1016/j.neuropsychologia.2017.03.018.

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.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci.: Off. J. Soc. Neurosci 14 (4), 1908–1919.

Zendel, B.R., Lagrois, M.-É., Robitaille, N., Peretz, I., 2015. Attending to pitch information inhibits processing of pitch information: the curious case of amusia. J. Neurosci.: Off. J. Soc. Neurosci 35 (9), 3815–3824. https://doi.org/10.1523/JNEUROSCI.3766-14.2015.

Zhang, C., Shao, J., Huang, X., 2017. Deficits of congenital amusia beyond pitch: evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. PLoS One 12 (8), e0183151. https://doi.org/10.1371/journal.pone.0183151.

SUPPLEMENTARY DATA

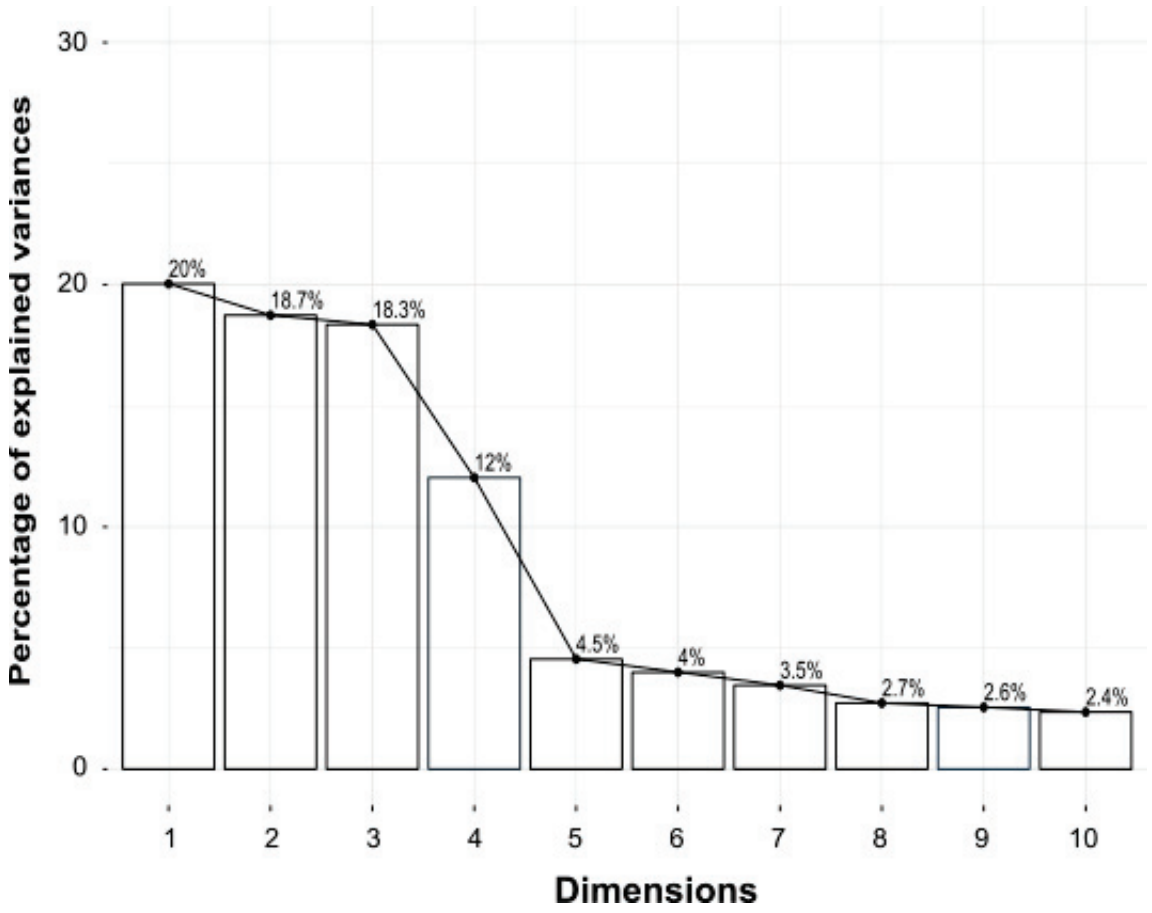

Figure A1: In the multifactorial analysis on categorization data for vowel stimuli*, four dimensions explain 69% of the data variance. We kept the first four dimensions for the rest of our analysis.*

	Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Correct categorization						
	Material + Group + Emotion + Material x Emotion	0,053	0,367	10,423	1,02E+13	2,717
	Material + Group + Material x Group + Emotion + Material x Emotion	0,053	0,194	4,345	5,41E+12	2,37
	Material + Group + Emotion + Material x Emotion + Group x Emotion	0,053	0,18	3,964	5,02E+12	3,517
	Material + Group + Material x Group + Emotion + Material x Emotion + Group x Emotion	0,053	0,113	2,292	3,14E+12	2,034
	Material + Group + Material x Group + Emotion + Material x Emotion + Group x Emotion + Material x Group x Emotion	0,053	0,076	1,486	2,12E+12	2,842
	Material + Emotion + Material x Emotion	0,053	0,069	1,33	1,91E+12	7,628
	Material + Group + Emotion	0,053	1,458E -4	0,003	4,06E+09	2,082
	Material + Group + Material x Group + Emotion	0,053	8,388E -5	0,002	2,33E+09	1,724
	Material + Group + Emotion + Group x Emotion	0,053	6,412E -5	0,001	1,78E+09	1,958
	Material + Group + Material x Group + Emotion + Group x Emotion	0,053	3,965E -5	7,14E-04	1,10E+09	2,217
	Material + Emotion	0,053	3,510E -5	6,32E-04	9,76E+08	1,081
	Material + Group	0,053	1,990E -6	3,58E-05	5,53E+07	2,49
	Material + Group + Material x Group	0,053	1,121E -6	2,02E-05	3,12E+07	5,79
	Material	0,053	5,382E -7	9,69E-06	1,50E+07	1,374
	Group + Emotion	0,053	2,895E -12	5,21E-11	80,495	1,419
	Group + Emotion + Group x Emotion	0,053	9,068E -13	1,63E-11	25,212	1,864
	Emotion	0,053	8,555E -13	1,54E-11	23,784	0,397
	Group	0,053	1,086E -13	1,95E-12	3,019	0,923
	Null model (incl, subject)	0,053	3,597E -14	6,47E-13	1	
Intensity						
	Material + Emotion + Material x Emotion	0,053	0,617	28,989	3,92E+29	1,852
	Group + Material + Emotion + Material x Emotion	0,053	0,3	7,719	1,91E+29	4,031
	Group + Material + Group x Material + Emotion + Material x Emotion	0,053	0,057	1,087	3,62E+28	2,872
	Group + Material + Emotion + Group x Emotion + Material x Emotion	0,053	0,021	0,394	1,36E+28	7,696
	Group + Material + Group x Material + Emotion + Group x Emotion + Material x Emotion	0,053	0,004	0,075	2,63E+27	2,648
	Group + Material + Group x Material + Emotion + Group x Emotion + Material x Emotion + Group x Material x Emotion	0,053	4,630E -4	0,008	2,95E+26	4,905

Material + Emotion	0,053	7,617E -17	1,37E-15	4,85E+13	2,562
Group + Material + Emotion	0,053	3,056E -17	5,50E-16	1,94E+13	7,748
Emotion	0,053	2,241E -17	4,03E-16	1,43E+13	0,818
Group + Emotion	0,053	8,597E -18	1,55E-16	5,47E+12	1,063
Group + Material + Group x Material + Emotion	0,053	5,619E -18	1,01E-16	3,57E+12	2,082
Group + Material + Emotion + Group x Emotion	0,053	1,728E -18	3,11E-17	1,10E+12	3,197
Group + Emotion + Group x Emotion	0,053	5,142E -19	9,26E-18	3,27E+11	1,816
Group + Material + Group x Material + Emotion + Group x Emotion	0,053	3,575E -19	6,44E-18	2,27E+11	4,237
Material	0,053	2,340E -30	4,21E-29	1,488	4,547
Null model (incl. subject)	0,053	1,572E -30	2,83E-29	1	
Group + Material	0,053	7,753E -31	1,40E-29	0,493	1,826
Group	0,053	5,391E -31	9,70E-30	0,343	1,268
Group + Material + Group x Material	0,053	1,551E -31	2,79E-30	0,099	2,893

Table A1: Results of the Bayesian mixed repeated measures ANOVAs on sentence and vowel stimuli (Material), for correct categorization scores and intensity ratings.

C. Study 2: Pre-attentive processing of neutral and emotional sounds in congenital amusia

Pre-attentive processing of neutral and emotional sounds in congenital amusia

Agathe Pralus[1,2], Marie Gomot[3], Jackson Graves[1,4], Fanny Cholvy[1,2], Lesly Fornoni[1,2], 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] UMR 1253, INSERM, Université de Tours, Tours, France

[4] Department of Psychology, University of Minnesota, Minneapolis

* Equally contributing authors.

Corresponding author: agathe.pralus@inserm.fr CRNL, 95 Bd Pinel 69675 Bron Cedex, France

Declarations of interest: none

Abstract

Congenital amusia is a life-long deficit of musical processing. In a previous behavioral study, we revealed that congenital amusics might be impaired for explicit emotional prosody recognition, but not for its implicit processing. With the aim to investigate amusics' automatic processing of prosody, the present study measured electroencephalography when amusic and control participants listened passively to neutral and emotional vowels presented within an oddball paradigm. The MMN was rather preserved for all deviants in amusia, whereas an earlier negative component was found decreased in amplitude in amusics compared to controls for the neutral and sadness deviants. For the most salient deviant (anger), the P3a was decreased in amplitude for amusics compared to controls. These results showed some preserved automatic detection of emotional deviance in amusia despite an early deficit to process subtle acoustic changes. In addition, the automatic attentional shift in response to salient deviants at later processing stages was reduced in amusics. Between-group differences were larger over bilateral prefrontal areas, previously shown to display functional impairments in congenital amusia. Our present study thus provides further understanding of the dichotomy between implicit and explicit processing in congenital amusia, in particular for vocal stimuli with emotional content.

Keywords: tone deafness, emotion, implicit processes, ERP, MMN, P3a, prosody

Introduction

Congenital amusia, also known as tone-deafness, is a life-long deficit of music processing. This deficit seems to affect one to two percent of the general population (Peretz et al., 2007; Peretz & Vuvan, 2017), with potentially genetic origins (Peretz et al., 2007). Individuals with congenital amusia show no hearing impairments or brain lesions that could explain their deficit. They are usually unable to sing in tune or detect an out-of-key note (see Peretz, 2016; Tillmann et al., 2015 for reviews). Several studies have revealed a specific pitch processing deficit in congenital amusia, with pitch perception tasks (Hyde & Peretz, 2004; Peretz et al., 2003) and pitch memory tasks (Albouy et al., 2016; Graves et al., 2019; Tillmann, Lévêque, et al., 2016; Williamson & Stewart, 2010). The pitch deficit was observed for non-musical material, such as isolated pitches or tone pairs (Albouy et al., 2016; Foxton et al., 2004; Peretz et al., 2009), as well as tone sequences or melodies (see Tillmann et al., 2015 for a review). As pitch processing is relevant beyond the musical domain, the investigation of congenital amusia has been extended to speech perception abilities. While some early studies did not report any deficit of speech processing in amusic individuals (Ayotte et al., 2002; Tillmann et al., 2009; Williamson & Stewart, 2010), more recent studies have revealed specific impairments of speech contour perception and intonation recognition in congenital amusia (Jiang et al., 2010; Liu et al., 2015, 2017; Nan et al., 2016; Nguyen et al., 2009; Tillmann, Burnham, et al., 2011; Tillmann, Rusconi, et al., 2011).

As pitch is essential to process emotions both in speech and music, some studies have started to investigate emotional processing in congenital amusia. Regarding musical emotion perception, congenital amusics have demonstrated either a mild impairment or no impairment in recognition tasks (Gosselin et al., 2015; Lévêque et al., 2018; Marin et al., 2015), but have shown preserved intensity ratings of the emotions (Lévêque et al., 2018). Regarding emotion perception in speech, referred to as emotional prosody, congenital amusics have demonstrated

a mild deficit of recognition (Lima et al., 2016; Pralus et al., 2019; Thompson et al., 2012), which was more pronounced for short vowels (with few acoustic cues) than long sentences (Lolli et al., 2015; Pralus et al., 2019). This recognition deficit was the largest for sadness stimuli (Pralus et al., 2019; Thompson et al., 2012), which tended to be confounded with neutral stimuli (Pralus et al., 2019). Interestingly, when congenital amusics were asked to rate the intensity of emotional prosody stimuli, they did not show any deficit compared to matched controls, even for vowels (Pralus et al., 2019). Intensity ratings of emotions can be given without precise categorical representation of the emotion or explicit labeling, suggesting some preserved implicit processing of emotions in amusia (Lévêque et al., 2018; Pralus et al., 2019). For music material, preserved implicit processing of pitch in amusia has been reported, even though explicit processing has been shown to be disrupted (Lévêque et al., 2018; Omigie et al., 2013; Pralus et al., 2019; Tillmann et al., 2012, 2014; Tillmann, Lalitte, et al., 2016). For instance, congenital amusics were able to process pitch changes as well as pitch incongruity (Peretz et al., 2009; Zendel et al., 2015), even though they were unable to detect these changes or incongruities when explicitly asked to do so (Moreau et al., 2009; Omigie et al., 2012; Tillmann, Lévêque, et al., 2016). This recent research suggests congenital amusia to be a disorder of consciousness related to pitch representations (Albouy et al., 2016; Marin et al., 2015; Moreau et al., 2009, 2013; Omigie et al., 2013; Peretz, 2016; Peretz et al., 2009; Stewart, 2011; Tillmann, Lalitte, et al., 2016).

Aiming to further investigate this hypothesis, in the present study, we recorded several electrophysiological measures that have been previously used to explore implicit processes in the typical and the pathological brain. One target measure, which reflects pre-attentional processes in the brain, is the well-studied Mismatch Negativity (MMN) (Näätänen et al., 2007; Näätänen & Alho, 1995). This negative ERP component emerges when a deviant event appears in a repetitive auditory sequence (referred to as the oddball paradigm). It is considered to be

automatic as it can be recorded even when participants are actively engaged in another task (Näätänen et al., 2007; Näätänen & Alho, 1995). The MMN signal is very robust, stable, and found in most control participants at the individual level (Chen et al., 2018; Kraus et al., 1992). It is generally reported to originate from supratemporal and frontal cortical regions (Näätänen et al., 2007; Näätänen & Alho, 1995). An MMN can be induced not only by simple acoustic deviants, as classically studied (Näätänen et al., 2007; Peretz et al., 2005), but also by emotional deviant events (Goydke et al., 2004). For emotional prosodic material, such as vowels, an MMN can be induced by an emotional deviant, compared to a neutral standard (Carminati et al., 2018; Charpentier et al., 2018). This emotional MMN occurs generally at a shorter latency and is larger than for neutral deviant (Schirmer et al., 2005, 2016).

For non-emotional material, amusics' automatic brain response to acoustic changes has been studied in passive listening paradigms with pitch tone deviants or tone-language stimuli (Fakche et al., 2018; Moreau et al., 2009, 2013; Nan et al., 2016; Omigie et al., 2013; Zhang & Shao, 2018). Using pitch change passive paradigms, amusics' early change-related evoked potentials, such as the MMN, were decreased in amplitude in comparison to controls for small pitch changes (Fakche et al., 2018; Moreau et al., 2009, 2013). When the pitch change was large enough (200 cents), the MMN seemed to be preserved in amusics (Moreau et al., 2009, 2013). Omigie et al. (2013) used real melodies to investigate amusics' and controls' brain responses as a function of the degree of expectedness of the notes (Omigie et al., 2013). The results revealed that with increased unexpectedness the early negativity (in the N1 latency range) increased for controls, but not for amusics. It suggests a deficit in the processing of musical structures at early processing stages, in keeping with the results of Albouy et al. (2013) in an active short-term memory task for melodies. When tone-language stimuli were used, amusics did not demonstrate any decrease of MMN in response to lexical tones (Nan et al., 2016; Zhang & Shao, 2018).

In active paradigms, similar result patterns regarding early brain response to different acoustic changes in congenital amusia have been observed (Braun et al., 2008; Peretz et al., 2005, 2009). For pitch change detection tasks, the MMN was decreased for amusics (compared to controls) only for small pitch deviants (Peretz et al., 2005, 2009). The replacement of the correct tone with an incorrect (out-of-tune) deviant tone at the end of familiar melodies revealed a decreased early negativity in amusics compared to controls (Braun et al., 2008). For language material, in particular intonation processing with statements and questions, the early negativity was preserved in amusics, but the N2 response was decreased in response to incongruent pairs of tones (Lu et al., 2015).

Some alterations in deviance detection have been observed also for later components, such as the P3 (Braun et al., 2008; Lu et al., 2015; Moreau et al., 2009, 2013; Peretz et al., 2009; Zhang & Shao, 2018). For pitch change detection tasks using tones, a decreased P3 was observed for amusics (in comparisons to controls) only for small pitch changes (25 cents), but not otherwise (Braun et al., 2008; Moreau et al., 2013). For lexical tone changes, smaller P3a and P3b were observed in amusics compared to controls for small lexical tone changes (high rising vs. low rising tone) (Zhang & Shao, 2018).

Overall, some results have shown decreased early electrophysiological markers related to pitch deviance detection in congenital amusia, mostly for small pitch changes, and sometime together with a reduction of the subsequent P3a. However, the pattern of automatic pitch processing in speech and music in congenital amusia still needs further investigation.

Our previous behavioral study investigating emotional prosody in congenital amusia has suggested preserved implicit prosody processing (Pralus et al., 2019). With the aim to further investigate amusics' automatic processing of prosody, the present study measured electroencephalography (EEG) when participants listened passively to vowels presented within an oddball paradigm. Emotionally neutral vowel served as the standard and either emotional

(anger and sadness) or neutral vowels as deviants. Evoked potentials were compared between participants with congenital amusia and control participants matched in age, education, and musical training. Emotional deviants (anger and sadness) and neutral deviant were chosen from the material of our previous study (Pralus et al., 2019) aiming for similar F0 difference compared to the neutral standard. Anger was the best recognized emotion by amusics whereas sadness was not well recognized and often confused with neutrality. Hence, these two emotional deviants had different patterns of recognition in the two participant groups, while intensity ratings were similar across groups for these stimuli. We hypothesized that early automatic processing of emotion deviancy will be impaired in amusics compared to controls, with potentially different responses to neutral and emotional deviancy in these two groups.

Material and Methods

Participants

Nineteen amusic participants and twenty-one control participants matched for gender, age, laterality, education, and musical training (as defined by years of instruction of an instrument) at the group level were included in the study (**Table 1**). They all gave written informed consent to participate in the experiment. Prior to the main experiment, all participants were tested with a subjective audiometry, the Montreal Battery of Evaluation of Amusia (Peretz et al., 2003) to diagnose amusia, and a Pitch Discrimination Threshold (PDT) test (Tillmann et al., 2009). A participant was considered amusic if he/she had a global MBEA score below 23 (maximum score = 30) and/or a MBEA pitch score (average of the first three subtests of the MBEA) inferior to 22 (maximum score = 30). All control participants had a global MBEA score above 24.5 and a MBEA pitch score above 23.3 (see **Table 1**). All participants had normal hearing (hearing loss inferior to 30 dB at any frequency in both ears). Study procedures were approved by a national ethics committee. Participants provided written informed consent prior to the experiment and were paid for their participation.

	Amusics (n=19)	**Controls (n=21)**	**p-value (group comparison)**
Age (years)	30.7 (±14.38) Min: 18 Max: 56	32.33 (±14.5) Min: 19 Max: 64	0.72
Education (years)	15 (±2.67) Min: 10	15.23 (±2.19) Min: 12	0.76

	Max: 20	Max: 20	
Musical training (years)	0	0.048 (±0.22) Min: 0 Max: 1	0.33
Sex	9M 10F	8M 13F	0.55
Handedness	5L 14R	4L 17R	0.58
MBEA score	22.02 (±1.8) Min: 16.83 Max: 24.5	26.45 (±1.04) Min: 24.8 Max: 28.5	**<0.001**
MBEA pitch score	21.05 (±1.97) Min: 15.67 Max: 23.67	26.6 (±1.42) Min: 23.33 Max: 28.67	**<0.001**
PDT (semitones)	1.33 (±1.48) Min: 0.11 Max: 4.99	0.29 (±0.15) Min: 0.08 Max: 0.71	**0.007**

Table 1: Characteristics of the participants in both groups. The MBEA (Montreal Battery for the Evaluation of Amusia, Peretz et al., 2003) score corresponds to the average of the six subtests of the battery (maximum score = 30, cut off: 23). Pitch mean score corresponds to the average of the three pitch subtests in the MBEA (scale, contour and interval, cut off: 22). Note that a participant was considered as amusic if any of these two measures (MBEA score, MBEA pitch score) was below the cut-off. PDT: Pitch Discrimination Threshold (see Tillmann et al., 2009). For each variable (except sex and handedness), the mean value in each group is reported along with the standard deviation in parentheses. Groups were compared with t.tests (two sided), except for sex and handedness where a Chi2 test was used (Qobs=0.35 and Qobs=0.3, respectively).

Stimuli

Four vowels /a/ were selected from a larger material set, all produced with female voices (Charpentier et al., 2018), and used in a previous behavioral study with amusic (N=18) and control (N=18) participants (Pralus et al., 2019). All stimuli lasted 400 ms and were equalized in RMS amplitude. The stimuli were selected based on their recognition scores in the behavioral task (Pralus et al., 2019, see **Table 2**) as follows: the neutral deviant and standard were equally well recognized by all participants; the anger deviant was selected as an easy deviant (equally well-recognized by both groups); the sadness deviant was selected as a difficult deviant for amusics. We added the constraint that all stimuli should be similar in pitch and should have received similar intensity ratings (for emotional stimuli) (see **Table 2** for details**)**. Acoustic parameters (pitch mean, spectral flux mean, brightness mean, roughness mean, inharmonicity mean, and attack time) of the stimuli were computed with the MIR toolbox (Lartillot & Toiviainen, 2007); **Table 2**). Each parameter (except Attack Time) was computed with a temporal frame of 50ms by default. We then computed the average of each parameter across time (see **Table 2**).

Acoustic parameters	**Neutral standard**	**Neutral deviant**	**Sadness deviant**	**Anger deviant**
Pitch mean (Hz)	241	199	228	278
Spectral flux mean (a.u.)	17.19	9.33	25.75	68.70
Brightness mean (a.u.)	0.20	0.13	0.23	0.27

Roughness mean (a.u.)	22.98	38.09	14.21	114.80
Inharmonicity mean (a.u.)	0.18	0.17	0.27	0.45
Attack time (s)	0.028	0.039	0.056	0.13
Behavioral data (Pralus et al., 2019)				
% Correct recognition in Controls	100	83	94	72
% Correct recognition in Amusics	94	83	56	67
Mean Intensity ratings in Controls	NA	NA	2.8	2.6
Mean Intensity ratings in Amusics	NA	NA	2.8	2.4

192 **Table 2**: **Acoustic parameters of the stimuli and associated behavioral data from Pralus et al.,**
193 ***(2019)***. The acoustic parameters were computed with the MIR Toolbox (Lartillot & Toiviainen, 2007),
194 with a temporal frame of 50ms. a.u.: arbitrary units. Percentage of correct emotion recognition and
195 intensity ratings (on a scale from 1 to 5) for these stimuli are from Pralus et al. *(2019)* and were obtained
196 from 18 congenital amusics and 18 matched controls. NA: not applicable, no intensity ratings was given
197 for neutral stimuli.

198

Procedure

The experiment took place in a sound-attenuated room. Participants watched a silent movie with subtitles, they were told to not pay attention to the sounds played over headphones. The recording session lasted 45 minutes.

EEG recordings and ERP measurements

The entire experimental paradigm was composed of three oddball blocks, each with one type of deviant (Neutral, Sadness, Anger) and one block with equiprobable stimuli. For each oddball block, 700 standards and 140 deviants were played. Two consecutive deviants were separated by at least three standards. During the equiprobable block, each of the 4 stimuli were played equally often (144 times each, 576 stimuli in total), with no more than two repetitions of the same stimulus in a row. The stimulus onset asynchrony (SOA) was always 700 ms.

EEG was recorded using 31 active electrodes (BrainAmp/Acticap, Brain Products, Germany) with a nose reference, with a sampling frequency of 1000Hz (bandwidth 0.016-1000 Hz). Eye movements were recorded with an electrode under the left eye (offline re-referenced to Fp1). ELAN software was used for EEG signal processing (Aguera et al., 2011). Band-stop filters centered around 50Hz and 150Hz were applied to the EEG signal to remove power line artifacts. Independent Component Analysis was performed on the EEG signal to remove artifacts due to eye movements and heartbeat (Delorme & Makeig, 2004). Averaging was done for each deviant and standard separately, in the three oddball blocks and the equiprobable block. Standards occurring after a deviant were not averaged. Averaging was done on a 700ms time-window (from -200 ms to 500 ms around stimulus onset). Trials with peak-to-peak amplitude variation exceeding 150 µV at any electrode were rejected. Noisy electrodes were interpolated. A 2-30Hz band-pass Butterworth filter (order 4) was applied to the evoked potentials. ERPs were baseline-corrected by subtracting the average of the signal in the 100ms before the stimulus. The difference wave for each type of deviant (Neutral, Sadness, Anger) was obtained by subtracting

the response to the deviant from the response to the standard in the same block of the oddball paradigm[1]. Grand-averaged curves were obtained for both groups (Amusics and Controls). The emergence of deviance-related ERPs (MMN and P3a in particular) was assessed with the comparison of deviant and standard ERPs using a nonparametric cluster-based permutation analysis (1000 permutations), in each group, for each of the three deviants. A first threshold of $p<0.05$ was used for permutation-based paired t-tests for each sample. Clusters were labeled as significant for $p<0.05$ at the end of the permutations, controlling for multiple comparisons in space (31 electrodes) and time. Based on the union of these emergence tests in both groups, two or three time windows of interest were selected for each emotion. For neutral deviant, three time-windows were selected: 67-130ms, 130-205ms, 225-310ms. For sadness deviant, three time-windows were selected: 77-140ms, 140-200ms, 220-295ms. For anger deviant, two time-windows were selected: 113-205ms, 217-299ms. The first window corresponds to an early negativity at the latency of the N1 (neutral and sadness deviant only), the next one to the MMN, and the last to the P3a.

Statistical analysis

Based on the emergence tests described above, a set of fronto-central electrodes was selected for the main analysis. Average amplitude for electrode sites along the antero-posterior axis (four levels) and for the two sides (pre-frontal=Fp1, Fp2, frontal=F3, F4, fronto-central=FC1, FC2, central=C3, C4, odd numbers correspond to electrodes on the left side, even numbers on the right side) were computed for each participant, for each type of deviant, in each of the time windows of interest. For each emotion (neutral, sadness, anger) and for each time-window (early negativity, MMN, P3, except for anger for which there was no early negativity), a

[1] The equiprobable stimuli could not be used as the reference stimulus to compute difference ERPs, as in the equiprobable block, the anger sound (and to a lesser extent the neutral deviant sound) elicited a negativity compared to the other equiprobable sounds in the latency range of the MMN, suggesting that MMNs were elicited within this sequence (see Figure S1).

Bayesian repeated-measures Analysis of Variance (ANOVA) was performed with group (Amusics, Controls) as a between-subjects factor, and localization (Fp, F, FC and C) and side (left, right) as within-subject factors[2].

We report Bayes Factor (BF) as a relative measure of evidence. To interpret the strength of evidence (according to Lee & Wagenmakers, 2014), we considered a BF under three as weak evidence, a BF between three and 10 as positive evidence, a BF between 10 and 100 as strong evidence and a BF higher than 100 as a decisive evidence. BF_{10} indicates the evidence of H1 (a given model) compared to H0 (the null model), and $BF_{inclusion}$ indicates the evidence of one effect over all models. As no post-hoc tests with correction for multiple comparison have as yet been developed for Bayesian statistics (Wagenmakers et al., 2017, 2018), we used t-tests with Holm-Bonferroni correction for multiple comparisons.

Data availability

Raw data were generated at Lyon Neuroscience Research Center (France). Derived data supporting the findings of this study are available from the corresponding author upon request.

Results

Based on the emergence tests, three deviance-related ERPs were identified in the difference curves (**Figures 1-3**): (1) an early negativity was observed, namely a negative fronto-central deflection in a time-window of ~70-140ms after the stimulus onset; (2) the MMN was identified as the negative fronto-central deflection in a time-window of ~140-200ms after the stimulus onset, associated with the typical polarity inversion at the mastoids; (3) the P3a was identified as the positive fronto-central deflection in a time-window of ~220-300ms after the stimulus onset, with a polarity inversion at the mastoids. See Figures 1-3A for averaged curves over

[2] We also computed Bayesian ANOVA on the amplitude at midline electrodes (Fz, Cz, Pz) for each time-window for each emotion. See Supplementary analysis for details.

fronto-central sites and **Figures 1-3B** for topographies. For the Anger deviant difference curve, only two emergence windows were retrieved, corresponding to the MMN and the P3a. For precise emergence windows for each emotion and each group, see **Figures 1-3C**. Overall, the morphology of deviance-related responses was slightly different across emotions. In particular, there were differences in the latencies of the ERPs across emotions, these latencies were similar between groups.

Only averaged curves with neutral and sadness deviants showed an early negativity on fronto-central electrodes (**Figures 1-2B**), which is at the latency of the N1: For the neutral deviant, the emergence was between 67 and 130ms, whereas it was later for the sadness deviant, between 77 and 140ms. As habituation of the N1 was visible on standards in the oddball blocks (**Figure S2**), this early negativity might mainly reflect the different degrees of habituation of the N1 between standards and deviants. This effect was possibly less pronounced for the anger deviant, which had a slower attack time (see **Table 2**), and resulted in later auditory ERPs (see in particular the delay in the P50 with respect to the other stimuli in the equiprobable block, **Figure S1**). At the MMN latency, a typical ERP was observed for the three types of deviants, emerging at different latencies for each emotion: 130-205ms for neutrality, 140-200ms for sadness, 113-205ms for anger. For anger, the MMN peak was larger than the other two for both groups. At the P3a latency, the three deviance-related ERPs had an emerging peak with different latencies for each emotion: 225-310ms for neutrality, 220-295 for sadness, 217-299ms for anger.

Based on these observations, and in particular, that ERPs and their latencies were not identical in the three emotions, the main analyses were performed separately by emotion and by component.

*Response to a neutral deviant (*Figure 1*)*

1. Early negativity

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization, Group, and the interaction between the two (BF10=1.2e+9). This model was 2.3 times better than the model with the main effect of Localization (BF10=5.3e+8), 3.6 times better than the model with the main effects of Localization and Group (BF10=3.37e+8), and 4.9 times better than the model with the main effects of Localization, Group, Side and the interaction between Group and Localization (BF10=2.44e+8). The best model was at least 11.5 times better than the other models (BF10<1.04e+8). This was confirmed by a decisive specific effect of Localization (BFinclusion=4.8e+8), a positive effect of the interaction between Localization and Group (BFinclusion=3.06) and no other specific effects (BFinclusion<1.07). According to post-hoc tests, amplitude at Fp sites was smaller than amplitudes at F, FC and C (all pcorr<0.025), amplitude at FC was higher than amplitudes at F and C (all pcorr<0.025). Amusics had a significantly smaller early negativity than controls. Specifically, amusics had smaller amplitude at Fp compared to amplitudes at F, FC, C (all pcorr<0.001), whereas no such pattern was observed in controls (all pcorr>0.39).

2. MMN

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=3.33e+14). This model was 1.3 times better than the model with the main effects of Localization and Group (BF10=2.52e+14), 7.1 times better than the model with the main effects of Localization and Side (BF10=4.66e+13), and 9.3 times better than the model with the main effects of Localization, Group, and Side (BF10=3.59e+13). The best model was at least 12 times better than the other models (BF10<2.75e+13). This was confirmed by a decisive specific effect of Localization (BFinclusion=6.43e+13), and no other

specific effects (BFinclusion<0.31). According to post-hoc tests, amplitude at Fp and C sites was smaller than amplitudes at F, FC (all pcorr<0.001). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls.

3. P3a

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization and Side (BF10=2.05e+10). This model was 1.9 times better than the model with the main effects of Localization, Side, and Group (BF10=1.07e+10), and 2.3 times better than the model with the main effects of Localization, Side, Group, and the interaction between Side and Group (BF10=8.86e+9). The best model was at least 10 times better than the other models (BF10<2.05e+9). This was confirmed by a decisive specific effect of Localization (BFinclusion=1.57e+9), a positive effect of Side (BFinclusion=6.01), and no other specific effects (BFinclusion<0.59). Amplitudes were larger over left side than right side. The Group effect emerging in the second best model showed that amusics tended to have a bigger P3a than controls. According to post-hoc tests, amplitudes at Fp and C sites were smaller than amplitudes at F and FC (all pcorr<0.001).

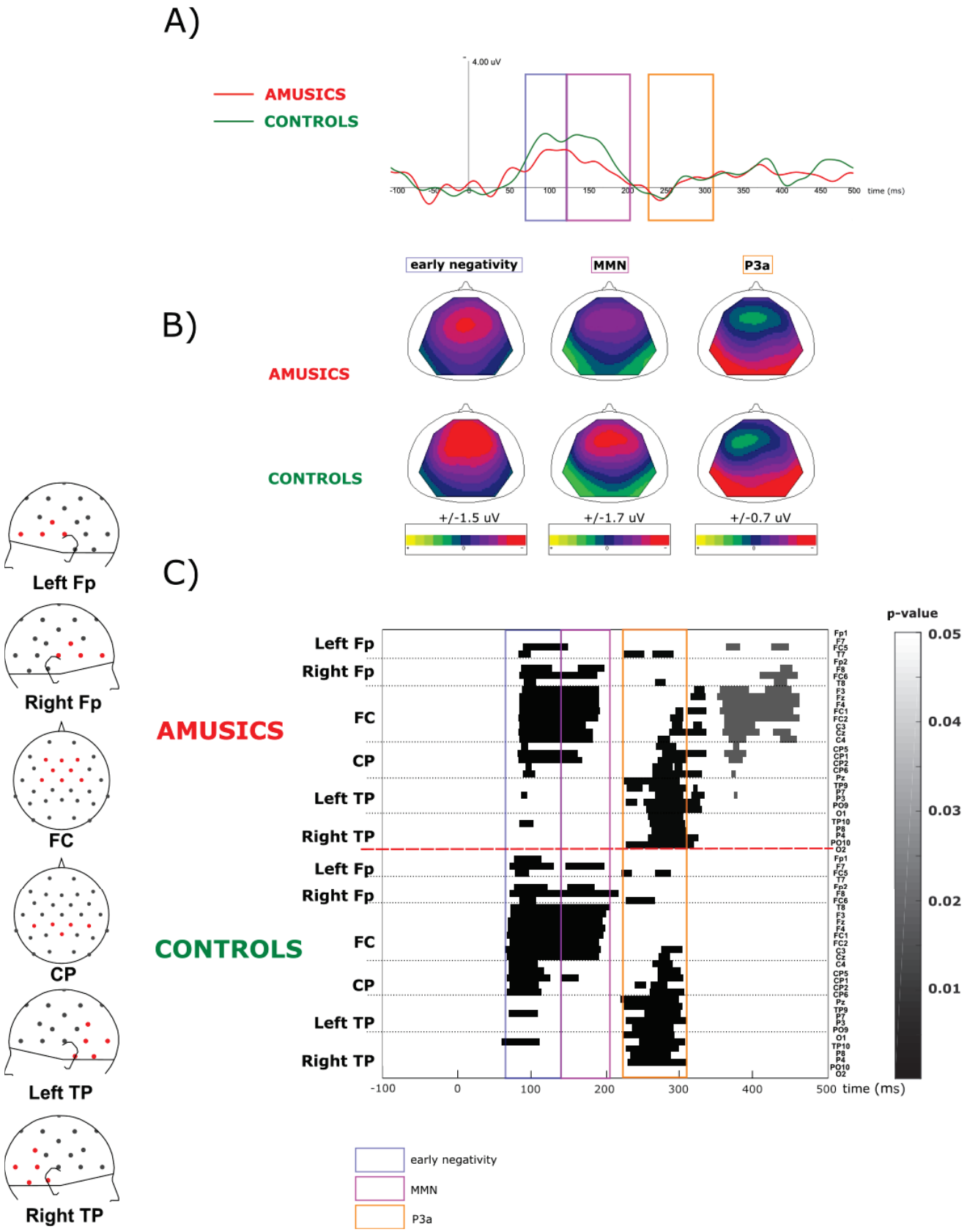

Figure 1: Evoked response to an emotionally neutral deviant in Amusics and Controls. A) Average curve of eight fronto-central electrodes (Fp1, Fp2, F3, F4, FC1, FC2, C3, C4) of the response to the neutral deviant minus the response to the neutral standard, for amusics and controls, negativity is up. **B)** Topographies for the three evoked potentials (early negativity, MMN and P3) over the emergence windows identified below, separately for amusics and controls. Amplitude scale is indicated for each ERP. **C)** Emergence of evoked responses for amusics and controls for each electrode, grouped by topography, emergence windows used for the analysis are in blue for early negativity (67-130ms), pink for the MMN (130-205ms), orange for the P3a (225-310ms). Fp= pre-frontal, FC=fronto-central, CP=centro-parietal, TP=temporo-parietal.

Response to a sadness deviant (Figure 2)

1. Early negativity

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization, Group and the interaction between the two (BF10=1.74e+9). This model was only 1.04 times better than the model with the main effect of Localization (BF10=1.68e+9), and 1.43 times better than the model with the main effects of Localization and Group (BF10=1.22e+9). The best model was at least 8.3 times better than the other models (BF10<2.1e+8). This was confirmed by a decisive specific effect of Localization (BFinclusion=9.36e+8), and no other specific effects (BFinclusion<1.37). Amusics had a smaller early negativity than controls. According to post-hoc tests, amplitude at Fp sites was smaller than amplitudes at F, FC (both pcorr<0.001), amplitudes at F and C were smaller than amplitude at FC (both pcorr<0.007). Specifically, amusics had smaller amplitude at Fp compared to amplitudes at F, FC, C (all pcorr<0.004), whereas controls had smaller amplitude at Fp compared only to FC (pcorr=0.021).

2. MMN

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=2.79e+10). This model was 1.64 times better than the model with the main effects of Localization and Group (BF10=1.7e+10), and 3.79 times better than the model with the main effects of Localization and Side (BF10=7.36e+9). The best model was at least 6.2 times better than the other models (BF10<4.52e+9). This was confirmed by a decisive specific effect of Localization (BFinclusion=1.2e+10), and no other specific effects (BFinclusion<0.29). The Group effect emerging in the second best model showed that the amusics tended to have a smaller MMN than controls. According to post-hoc tests, amplitudes at Fp and C were smaller than amplitudes at FC and F (all pcorr<0.004).

3. P3a

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=6.25e+15). This model was 1.63 times better than the model with the main effects of Localization and Group (BF10=3.83e+15), and 7.98 times better than the model with the main effects of Localization and Side (BF10=7.83e+14). The best model was at least 14 times better than the other models (BF10<4.44e+14). This was confirmed by a decisive specific effect of Localization (BFinclusion=∞), and no other specific effects (BFinclusion<0.23). The Group effect emerging in the second best model showed that amusics tended to have a smaller P3a than controls. According to post-hoc, amplitude at Fp was smaller than amplitudes at F, C and FC (all pcorr<0.015), amplitude at C was smaller than amplitudes at F and FC (both pcorr<0.041).

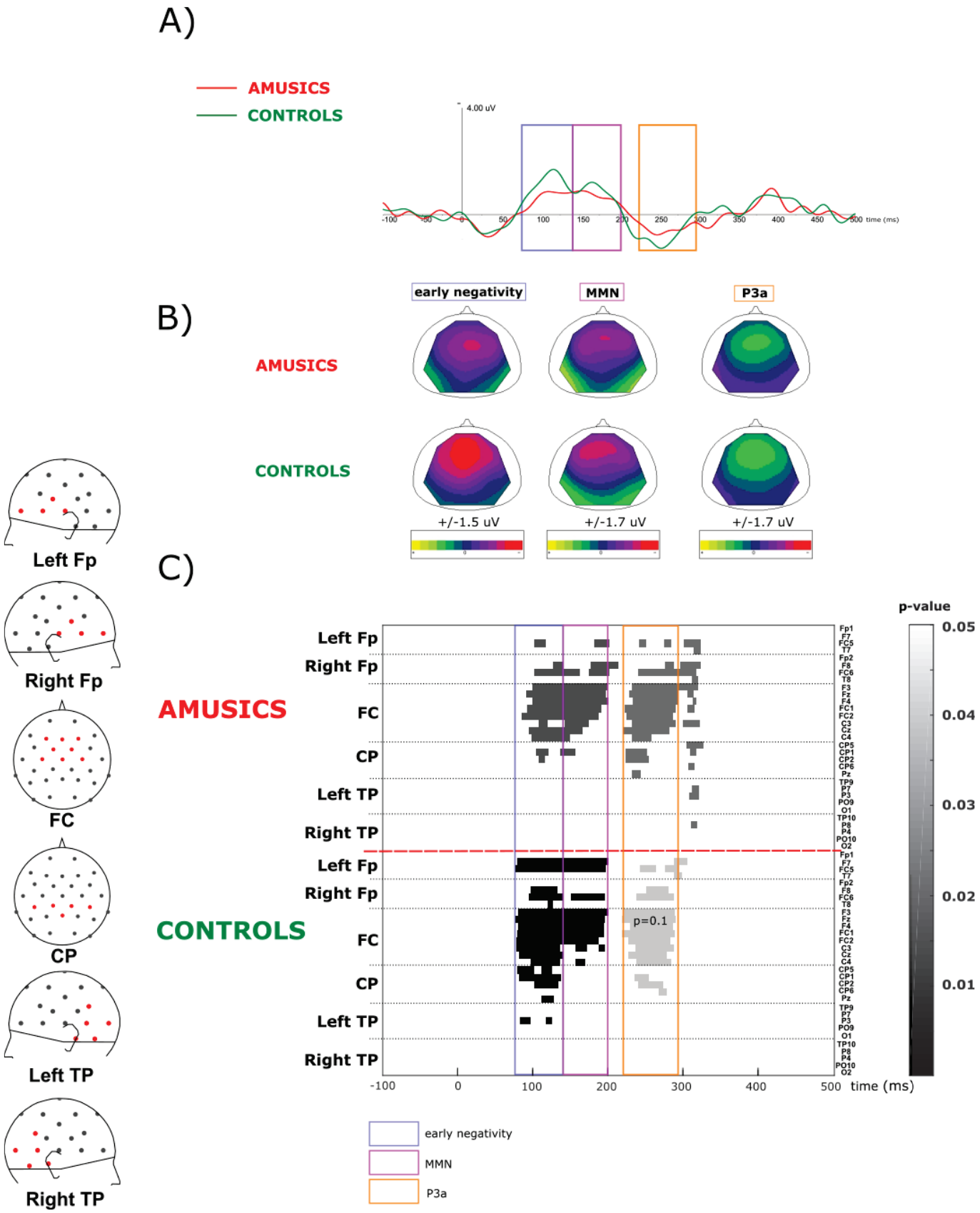

Figure 2: Evoked response to an emotional sadness deviant in Amusics and Controls. A) Average curve of eight fronto-central electrodes (Fp1, Fp2, F3, F4, FC1, FC2, C3, C4) of the response to the sadness deviant minus the response to the neutral standard, for amusics and controls, negativity is up. **B)** Topographies for the three evoked potentials (early negativity, MMN and P3) over the emergence windows identified below, separately for amusics and controls. Amplitude scale is indicated for each ERP. **C)** Emergence of evoked responses for amusics and controls for each electrode, grouped by topography, emergence windows used for the analysis are in blue for early negativity (77-140ms), pink for the MMN (140-200ms), orange for the P3a (220-295ms). The P3a only emerged in the amusic group at a pvalue of 0.05 for the permutation test, if the pvalue was set at 0.1 it also emerged in the control group. Fp= pre-frontal, FC=fronto-central, CP=centro-parietal, TP=temporo-parietal.

*Response to an anger deviant (*Figure 3*)*

1. MMN

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=5.99e+14). This model was 1.74 times better than the model with the main effects of Localization and Group (BF10=3.45e+14), and 4.68 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=1.28e+14). The best model was at least 8.14 times better than the other models (BF10<7.36e+14). This was confirmed by a decisive specific effect of Localization (BFinclusion=9.46e+13), and no other specific effects (BFinclusion<0.3). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls. According to post-hoc, amplitude at Fp sites was smaller than amplitudes at C, FC and F (all pcorr<0.006), amplitude at C and F were smaller than amplitude at FC (both pcorr<0.035).

2. P3a

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization and Group and the interaction between the two (BF10=3.25e+16). This model was 4.6 times better than the model with the main effects of Localization, Side, Group, and the interaction between Localization and Group (BF10=7.02e+15), and 6.1 times better than the model with the main effects of Localization and Group (BF10=5.3e+15). The best model was at least 11.9 times better than the other models (BF10<2.95e+15). This was confirmed by a decisive specific effect of Localization (BFinclusion=∞), and positive effects of Group (BFinclusion=4.99) and the interaction between Localization and Group (BFinclusion=9.23), and no other specific effects (BFinclusion<0.09). According to post-hoc tests, amplitude at Fp was smaller than amplitudes at C, FC and F (all pcorr<0.017), amplitudes at C and F were smaller than amplitude at FC (both pcorr<0.001), amplitude at C was smaller than amplitude at F (pcorr=0.032). Amusics had a significantly

smaller P3 compared to Controls. This group difference was especially observed at Fp sites (pcorr=0.081). Specifically, amusics had smaller amplitude at Fp compared to amplitudes at F, FC, C (all pcorr<0.001), whereas controls had smaller amplitude at Fp compared only to FC (pcorr=0.002).

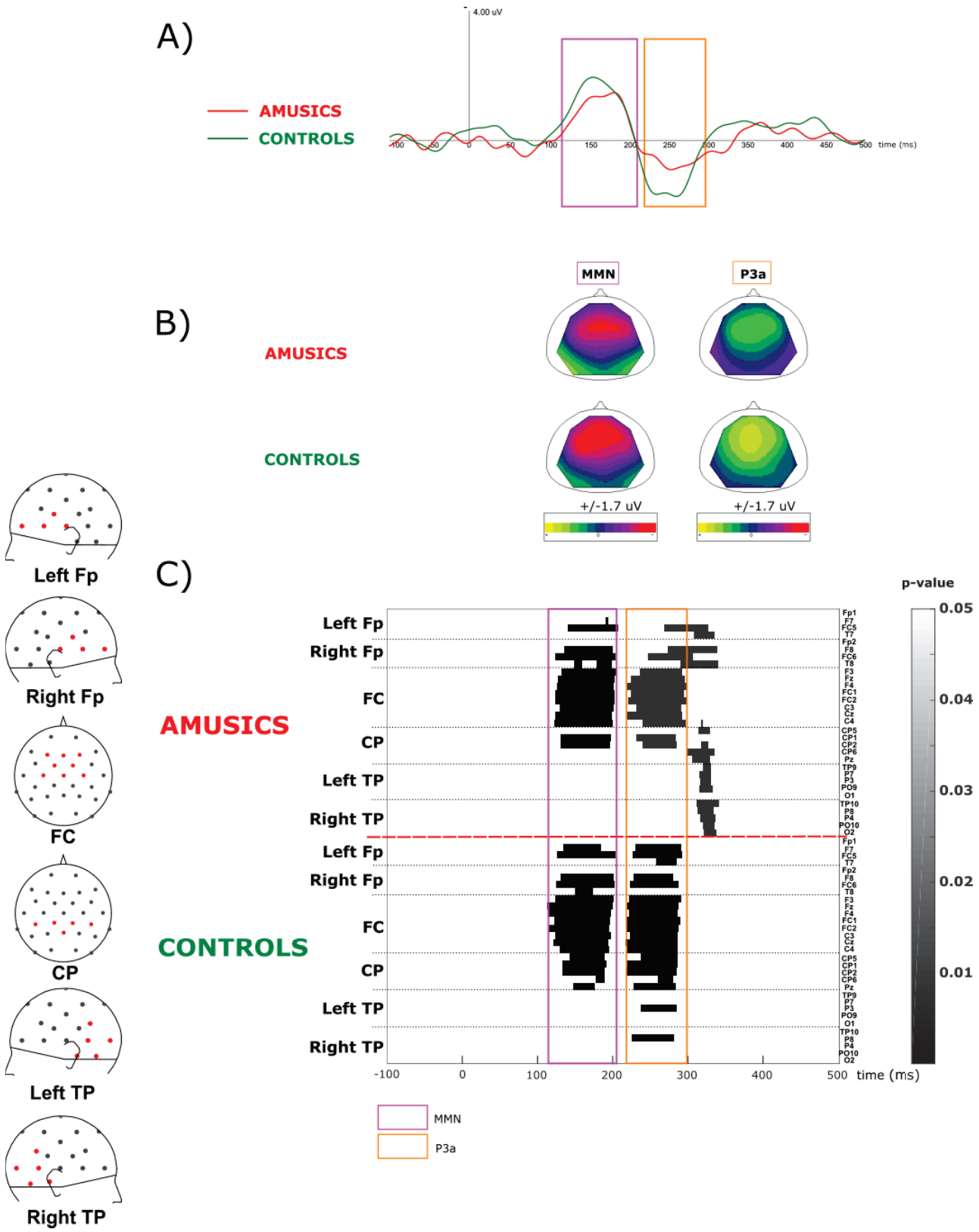

Figure 3: Evoked response to an emotional anger deviant in Amusics and Controls. A) Average curve of eight fronto-central electrodes (Fp1, Fp2, F3, F4, FC1, FC2, C3, C4) of the response to the anger deviant minus the response to the neutral standard, for amusics and controls, negativity is up. **B)** Topographies for the two evoked potentials (MMN and P3) over the emergence windows identified below, separately for amusics and controls. Amplitude scale is indicated for each ERP. **C)** Emergence of evoked responses for amusics and controls for each electrode, grouped by topography, emergence windows used for the analysis are in pink for the MMN (113-205ms), orange for the P3a (217-299ms). Fp= pre-frontal, FC=fronto-central, CP=centro-parietal, TP=temporo-parietal.

Comparisons between deviants

To investigate potential differences across emotions, we ran a Bayesian ANOVA with the additional within-subjects factor Emotion for each evoked potential.

1. Early negativity

This analysis included only the neutral and sadness deviants. After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization and Emotion (BF10=3.2e+10). This model was 1.4 times better than the model with the main effects of Localization, Emotion, Group and the interaction between Localization and Group (BF10=2.3e+10), and 1.6 times better than the model with the main effects of Localization, Emotion and Group (BF10=2.05e+10). The best model was at least 4.1 times better than the other models (BF10<7.8e+9). This was confirmed by a decisive specific effect of Localization (BFinclusion=6.17e+8), a small positive effect of Emotion (BFinclusion=2.68) and no other specific effects (BFinclusion<0.51). According to post-hoc tests, the amplitude at Fp sites was smaller than amplitudes at F, FC and C (all pcorr<0.012), amplitude at FC was larger than amplitudes at F and C (all pcorr<0.001). The early negativity for sadness was smaller than the one for neutrality. Amusics tended to have a smaller early negativity than controls. Specifically, amusics tended to have smaller amplitude at Fp compared to amplitudes at F, FC, C (all pcorr<0.001), whereas controls tended to have smaller amplitudes at Fp compared to amplitudes at FC only (pcorr=0.019).

2. MMN

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization and Emotion (BF10=1.6e+18). This model was 2.2 times better than the model with the main effects of Localization, Emotion and Group (BF10=7.15e+17), and 13.3 times better than the model with the main effects of Localization, Emotion and Side (BF10=1.2e+17). The best model was at least 18.8 times better than the other

models (BF10<8.49e+16). This was confirmed by a decisive specific effect of Localization (BFinclusion=2.82e+13) and Emotion (BFinclusion=782), and no other specific effects (BFinclusion<0.071). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls. According to post-hoc tests, amplitude at Fp and C sites was smaller than amplitudes at F, FC (all pcorr<0.001), amplitude at Fp was smaller than amplitude at C (pcorr=0.029). The MMN tended to be smaller for sadness compared to anger (pcorr=0.2).

3. P3a

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization, Emotion, Side, and Group, and the interaction between Emotion and Localization, between Emotion and Side, between Side and Localization, between Emotion and Group, and the triple interaction between Emotion, Localization and Side (BF10=9.85e+82). This model was 16.2 times better than the model with the main effects of Localization, Emotion, Side and Group and the interaction between Emotion and Localization, between Emotion and Side, between Side and Localization, between Emotion and Group, between Side and Group, and the triple interaction between Emotion, Localization and Side (BF10=6.08e+81), and 78 times better than the model with the main effects Localization, Emotion, Side and Group and the interaction between Emotion and Localization, between Emotion and Side, between Side and Localization, between Emotion and Group, between Localization and Group, and the triple interaction between Emotion, Localization and Side (BF10=1.26e+81). The best model was at least 99 times better than the other models (BF10<9.9e+80). This was confirmed by a decisive specific effect of Localization (BFinclusion=∞), Emotion (BFinclusion=∞), Side (BFinclusion=1164), Group (BFinclusion=6949), the interaction between Emotion and Localization (BFinclusion=1.24e+10), the interaction between Emotion and Side (BFinclusion=664), the

interaction between Emotion and Group (BFinclusion=33214), the interaction between Localization and Side (BFinclusion=676) and the interaction between Emotion, Localization and Side (BFinclusion=4740), and no other specific effects (BFinclusion<0.07). Amusics had a smaller P3a compared to Controls. According to post-hoc tests, amplitude at Fp was smaller than amplitudes at C, FC and F (all pcorr<0.016), amplitudes at C and F were smaller than amplitude at FC (both pcorr<0.016), amplitude at C was smaller than amplitude at F (pcorr<0.001). The P3a was larger for anger compared to sadness and neutrality (both pcorr<0.046), and larger for sadness compared to neutrality (pcorr<0.001). More specifically, amplitudes for neutrality was smaller than amplitude for anger and sadness at C, F and FC (all pcorr<0.002), amplitude for neutrality was smaller than amplitude for anger at Fp (pcorr=0.002). Neutrality had smaller amplitude than had sadness and anger for both left and right sides (all pcorr<0.001). Amplitudes were larger over left side than right side. This was driven by a smaller amplitude at C for right side compared to left side (pcorr=0.098). This difference was driven in particular by the difference between the two groups for anger (pcorr=0.1). In amusics, neutrality had smaller amplitude than had anger (pcorr=0.025), whereas in controls, neutrality had smaller amplitude than had anger and sadness (both pcorr<0.004).

The analyses of components amplitude at midline electrodes (Fz, Cz, Pz) are reported in the supplementary material. Only limited group effects were observed in these analyses, in keeping with the results reported above which reveal that between-group differences were mostly observed at prefrontal sites, for which we did not have a midline electrode in our 32-electrode montage. These results at midline electrodes further emphasize that the early negativity peaking at Fz and Cz was slightly more central than the MMN, which peaked at Fz. This is in agreement with the hypothesis that the early negativity included N1 refractoriness effects.

Discussion

Using an oddball paradigm with emotional prosody stimuli, we revealed the automatic brain responses of congenital amusic individuals compared to matched control participants for neutral and emotional verbal sounds. Based on previous behavioral and ERP results, we expected a decreased early automatic processing of deviancy in amusics compared to controls, with potentially different responses to neutral and emotional deviancy in these two groups. Amusics had reduced automatic processing of a neutral deviant compared to controls, with a diminished early negativity at the latency of N1 and a slightly reduced MMN. Similarly, the early processing of emotional stimuli (reflected by the early negativity at the latency of N1) was decreased in amusics compared to controls, yet with only slightly reduced emotional MMNs. The later P3a observed in response to a salient emotional deviant (anger) was strongly decreased in amusics compared to controls. These results suggest a differential processing of neutrality and emotions, with impaired pre-attentive processing of both neutral and emotional sounds in congenital amusia, at early cortical processing stages (around 100 ms) and in late processing stages associated with high-level cognitive processes (around 300 ms). The rather preserved MMN in between these altered processing stages suggest that change detection mechanisms can operate on degraded initial sound representations, at least in the case of large enough sound deviances.

Even if congenital amusia was first described to be music-specific (Ayotte et al., 2002; Peretz et al., 2003), recent evidence suggest that the pitch deficit in congenital amusia could also extend to speech material, even though to a lesser extent (Nguyen et al., 2009; Tillmann, Burnham, et al., 2011; Tillmann, Rusconi, et al., 2011; Zhang et al., 2017). In relation with the present study, congenital amusia is not only a music perception deficit but also a language processing deficit, in particular for non-verbal auditory cues such as emotional prosody (Lolli et al., 2015; Nguyen et al., 2009; Patel et al., 2008; Pralus et al., 2019; Thompson et al., 2012).

Impaired early encoding of auditory stimuli in congenital amusia

A smaller early negativity was observed in amusics compared to controls for neutral and sadness deviants. It points to amusics' increased difficulties to automatically process the deviants at early processing stages. This early negativity seems to correspond to N1 adaptation effects as the adaptation observed here occurs in the latency range of the N1, with a slightly different topography than the subsequent MMN.

In agreement with previous research (Albouy et al., 2013; Omigie et al., 2013), our results thus reveal an early deficit of auditory encoding in the amusics' brain. This early processing seems to be particularly less efficient for neutral stimuli, but can also be altered for emotional stimuli, as revealed by the results with the sadness deviant. Interestingly, taken together, the results suggest a general decrease of the early negativity in congenital amusia, observed both in the processing of pitch sequences (Albouy et al., 2013; Omigie et al., 2013) and in oddball contexts (current results). As suggested by stimulus-specific adaptation research (Carbajal & Malmierca, 2018; Malmierca et al., 2014; Pérez-González & Malmierca, 2014), a precise representation of the standard is necessary to elicit a strong N1 when the deviant is presented. However, if the representation of the standard is not precise, as in congenital amusics, the N1 elicited by the deviant remains similar to the N1 elicited by the standard, as revealed here using an oddball paradigm.

Interestingly, similar pitch deviance was used with the three types of deviant (the smaller pitch deviance was for the sadness deviancy). Even though we tried to match the acoustic differences between the three types of deviant as closely as possible, other acoustic features than pitch differentiated between the three emotions. These variations of acoustic parameters could explain at least in part the pattern of evoked responses in the two groups. As roughness and inharmonicity were higher for the anger deviant compared to sadness and neutrality, it could have helped amusics to correctly process this anger deviant and recognize it behaviorally

(Pralus et al., 2019). Indeed, previous reports suggest that amusics' emotional judgments are based largely on roughness and tempo rather than harmonicity cues, which are mostly used by controls (Gosselin et al., 2015; Lévêque et al., 2018; Marin et al., 2015). Moreover, the anger deviant was characterized by a longer attack time, in particular when compared to the neutral standard. This could explain why the pattern of the first evoked potentials in response to this deviant was different compared to the two other emotions, with no early negativity at the latency of N1. However, this specific pattern of responses was similar in the two groups.

Preserved automatic change detection and implicit processes in amusia

To investigate automatic change detection processes in congenital amusia, we studied the MMN evoked by the three deviants. As expected (Carminati et al., 2018; Charpentier et al., 2018), a MMN was induced by both emotional and neutral deviants, compared to a neutral standard in both groups. The MMN was larger for emotional deviants than for the neutral deviant (Schirmer et al., 2005, 2016).

No clear deficit of the MMN for the neutral or emotional deviants was observed in amusics compared to controls, suggesting at least a partially preserved automatic processing of emotional prosody in amusics, as previously shown with behavioral data (Lima et al., 2016; Lolli et al., 2015; Pralus et al., 2019; Thompson et al., 2012). This result is in line with previous research on automatic pitch processing in congenital amusia demonstrating only a small deficit of the congenital amusics' MMN for small pitch changes in tone sequences (Fakche et al., 2018; Moreau et al., 2009, 2013; Nan et al., 2016; Omigie et al., 2013; Zhang & Shao, 2018). It suggests that, despite an impaired early processing of the deviant, congenital amusics' brain is still able to automatically detect the change. However, even if it was not significant, we did observe a small decrease of the MMN to the neutral deviant in amusics, suggesting that this implicit knowledge in amusics' brain might not be fully sufficient in some cases to allow the change detection mechanisms underlying the MMN to produce as large error signals as in

controls. It is widely admitted that a correct sensory memory representation of the standard is needed to elicit an MMN (Näätänen et al., 2005). This would suggest that in congenital amusic participants, this memory representation is not as accurate as in controls. Thus, these results would contribute to the understanding of the deficit in congenital amusia as previously demonstrated with short-term memory tasks (Albouy et al., 2013, 2016; Fakche et al., 2018; Graves et al., 2019; Tillmann et al., 2009; Williamson & Stewart, 2010).

In combination with the analysis of the early negativity, these results show that acoustic sensitivity is impaired in congenital amusia, and do not seem to depend on emotional content of the stimulus. However, the more cognitive and memory-related comparison reflected by the MMN (Maess et al., 2007) seems to be less impaired in congenital amusia. In particular, it appears that this component would be only minimally impacted in congenital amusia when an emotional component is present in the stimulus. Such preserved automatic cortical processing steps could be the basis of the preserved implicit processes observed behaviorally in musical and emotional judgements (Lévêque et al., 2018; Pfeuty & Peretz, 2010; Pralus et al., 2019; Stewart, 2011; Tillmann et al., 2007; Tillmann, Lalitte, et al., 2016).

Decreased awareness of emotional stimuli in congenital amusia

To further investigate the potential deficit of awareness in congenital amusia for emotional stimuli, we analyzed the P3a in response to the three types of deviants. This ERP was larger for the emotional deviants, especially for anger, but was still detectable for the two deviants in the two participant groups. For the anger deviant, which elicited the largest P3a in controls, the P3a was strongly decreased in amusics compared to controls. A reduced P3a in amusics was previously shown with lexical tones (Zhang & Shao, 2018) and using tasks with small pitch changes in tone sequences (Braun et al., 2008; Moreau et al., 2009, 2013). The decreased P3a relates to an awareness deficit suggested in congenital amusia (Peretz et al., 2009), in particular for emotional stimuli (Lévêque et al., 2018; Pralus et al., 2019). Specifically, P3a is considered

to reflect automatic attentional orientation toward a salient deviant (Escera et al., 1998; Polich & Criado, 2006). Thus, congenital amusics would have a deficit to process unexpected novel sounds. However, amusics were still able to perform the recognition task for the anger deviant. These results suggest that when the automatic preattentional processes of the amusics reach a sufficient level (a sizeable MMN and a detectable P3a), they can perform the recognition task, despite this deficit at these late processing stages.

Brain networks involved in emotional prosody perception in congenital amusia

The group differences were mostly visible on bilateral pre-frontal electrodes. Interestingly, in congenital amusia, frontal regions were found to be altered (Albouy et al., 2013, 2019; Hyde et al., 2006, 2007, 2011). In particular, decreased gray and white matter volume of the inferior frontal cortices was observed in congenital amusia (Albouy et al., 2013, 2019; Hyde et al., 2006, 2007, 2011). As these regions are involved in emotional prosody processing (Frühholz et al., 2012; Liu et al., 2015), it could have been expected that amusics would have a deficit to perceive emotional prosody. However, our results with the MMN suggest a partial preservation of these circuits to automatically detect emotional prosody in congenital amusics. These results are in line with previous reports showing that the perception of emotional prosody does not only involve a fronto-temporal network, but also extend to other regions, such as probably the amygdala that detects salience and meaningful information (Frühholz et al., 2016), which would be preserved in congenital amusia. Further research using brain imaging (with fMRI for example) should investigate the brain networks involved in emotional perception in congenital amusia.

Conclusion

Our present findings shed new light on different aspects of automatic sound processing in congenital amusia, in particular for speech material and its emotional features. The observed impairments might lead to difficulties to process speech correctly in some situations. For

instance, in degraded conditions such as hearing in noise, challenging conditions for speech comprehension (Liu et al., 2015; Oxenham, 2008, 2012; Tang et al., 2018), amusics could have more difficulties to understand the speaker's emotions and intentions (Mcdonald & Stewart, 2008; Omigie et al., 2012). Moreover, this study gives further insight about the dissociation of implicit and explicit processing in congenital amusia (Lévêque et al., 2018; Omigie et al., 2013; Pralus et al., 2019; Tillmann et al., 2012, 2014; Tillmann, Lalitte, et al., 2016). It reveals the overall pattern of emotional perception in congenital amusia, from the first steps of cortical processing (Albouy et al., 2013; Omigie et al., 2013) to the late processing stages (P300, Moreau et al., 2009, 2013; Peretz et al., 2009), via the intermediate stage of change detection reflected by the MMN. This relatively preserved MMN might relate to preserved implicit processing in congenital amusia for music and emotional prosody stimuli.

Fundings

This work was conducted in 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).

References

Aguera, P.-E., Jerbi, K., Caclin, A., & Bertrand, O. (2011). ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. *Computational Intelligence and Neuroscience*, *2011*, 158970. https://doi.org/10.1155/2011/158970

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*. https://doi.org/10.3389/fnins.2019.01165

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports*, *6*, 18861. https://doi.org/10.1038/srep18861

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: A Journal of Neurology*, *136*(Pt 5), 1639–1661. https://doi.org/10.1093/brain/awt082

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusiaA group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. https://doi.org/10.1093/brain/awf028

Braun, A., McArdle, J., Jones, J., Nechaev, V., Zalewski, C., Brewer, C., & Drayna, D. (2008). Tune Deafness: Processing Melodic Errors Outside of Conscious Awareness as Reflected by Components of the Auditory ERP. *PLoS ONE*, *3*(6). https://doi.org/10.1371/journal.pone.0002349

Carbajal, G. V., & Malmierca, M. S. (2018). The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518784822

Carminati, M., Fiori-Duharcourt, N., & Isel, F. (2018). Neurophysiological differentiation between preattentive and attentive processing of emotional expressions on French vowels. *Biological Psychology*, *132*, 55–63. https://doi.org/10.1016/j.biopsycho.2017.10.013

Charpentier, J., Kovarski, K., Roux, S., Houy-Durand, E., Saby, A., Bonnet-Brilhault, F., Latinus, M., & Gomot, M. (2018). Brain mechanisms involved in angry prosody change detection in school-age children and adults, revealed by electrophysiology. *Cognitive, Affective & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-018-0602-8

Chen, C., Chan, C.-W., & Cheng, Y. (2018). Test–Retest Reliability of Mismatch Negativity (MMN) to Emotional Voices. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00453

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Fakche, C., Lecaignard, F., Lévêque, Y., Fornoni, L., Tillmann, B., & Caclin, A. (2018). Impaired short-term memory for pitch in congenital amusia: a MEG/EEG study of Mismatch Negativity. *Poster Presented at MMN2018: 8th Mis-match Neg-at-iv-ity Conference, Helsinki*.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness.' *Brain*, *127*(4), 801–810. https://doi.org/10.1093/brain/awh105

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific Brain Networks during Explicit and Implicit Decoding of Emotional Prosody. *Cerebral Cortex*, *22*(5), 1107–1117. https://doi.org/10.1093/cercor/bhr184

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *71*, 171–182. https://doi.org/10.1016/j.cortex.2015.06.022

Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, *21*(3), 351–359. https://doi.org/10.1016/j.cogbrainres.2004.06.009

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition*, *136*, 103614. https://doi.org/10.1016/j.bandc.2019.103614

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. *Journal of Neuroscience*, *27*(47), 13028–13032. https://doi.org/10.1523/JNEUROSCI.3039-07.2007

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

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain: A Journal of Neurology*, *129*(Pt 10), 2562–2570. https://doi.org/10.1093/brain/awl204

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 (New York, N.Y.: 1991)*, *21*(2), 292–299. https://doi.org/10.1093/cercor/bhq094

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*(9), 2630–2639. https://doi.org/10.1016/j.neuropsychologia.2010.05.009

Kraus, N., McGee, T., Sharma, A., Carrell, T., & Nicol, T. (1992). Mismatch negativity event-related potential elicited by speech stimuli. *Ear and Hearing*, *13*(3), 158–164.

Lartillot, O., & Toiviainen, P. (2007). Mir in Matlab (ii): A Toolbox for Musical Feature Extraction from Audio. *Austrian Computer Society*.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge Core - Psychology Research Methods and Statistics. https://doi.org/10.1017/CBO9781139087759

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*. https://doi.org/10.1037/neu0000461

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*. https://doi.org/10.1038/srep34911

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., & Wong, P. C. M. (2017). Perceptual learning of pitch direction in congenital amusia: evidence from Chinese speakers. *Music Perception*, *34*(3), 335–351.

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia*, *66*, 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001

Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: results from congenital amusia. *Frontiers in Psychology*, *6*, 1340. https://doi.org/10.3389/fpsyg.2015.01340

Lu, X., Ho, H. T., Liu, F., Wu, D., & Thompson, W. F. (2015). Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: an ERP study. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00385

Maess, B., Jacobsen, T., Schröger, E., & Friederici, A. D. (2007). Localizing pre-attentive auditory memory-based comparison: Magnetic mismatch negativity to pitch change. *NeuroImage*, *37*(2), 561–571. https://doi.org/10.1016/j.neuroimage.2007.05.040

Malmierca, M. S., Sanchez-Vives, M. V., Escera, C., & Bendixen, A. (2014). Neuronal adaptation, novelty detection and regularity encoding in audition. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00111

Marin, M. M., Thompson, W. F., Gingras, B., & Stewart, L. (2015). Affective evaluation of simultaneous tone combinations in congenital amusia. *Neuropsychologia*, *78*, 207–220. https://doi.org/10.1016/j.neuropsychologia.2015.10.004

Mcdonald, C., & Stewart, L. (2008). Uses and functions of music in congenital amusia. *Music Perception 25(4):345*. https://doi.org/10.1525/MP.2008.25.4.345

Moreau, P., Jolicoeur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 191–194. https://doi.org/10.1111/j.1749-6632.2009.04775.x

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain and Cognition*, *81*(3), 337–344. https://doi.org/10.1016/j.bandc.2013.01.004

Näätänen, R., & Alho, K. (1995). Generators of electrical and magnetic mismatch responses in humans. *Brain Topography*, *7*(4), 315–320. https://doi.org/10.1007/BF01195257

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*, *42*(1), 25–32.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nan, Y., Huang, W. T., Wang, W. J., Liu, C., & Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (MMN) among Mandarin speakers with congenital amusia. *Biological Psychology*, *113*, 59–67. https://doi.org/10.1016/j.biopsycho.2015.11.010

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 490–493. https://doi.org/10.1111/j.1749-6632.2009.04855.x

Omigie, D., Müllensiefen, D., & Stewart, L. (2012). The Experience of Music in Congenital Amusia. *Music Perception: An Interdisciplinary Journal*, *30*(1), 1–18. https://doi.org/10.1525/mp.2012.30.1.1

Omigie, D., Pearce, M. T., Williamson, V. J., & Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia*, *51*(9), 1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010

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

Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(39), 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech Intonation Perception Deficits in Musical Tone Deafness (congenital Amusia). *Music Perception: An Interdisciplinary Journal*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

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., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain: A Journal of Neurology*, *132*(Pt 5), 1277–1286. https://doi.org/10.1093/brain/awp055

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, *58*(3), 478–482. https://doi.org/10.1002/ana.20606

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*, 58–75.

Peretz, I., Cummings, S., & Dubé, M.-P. (2007). The Genetics of Congenital Amusia (Tone Deafness): A Family-Aggregation Study. *American Journal of Human Genetics*, *81*(3), 582–588.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics: EJHG*, *25*(5), 625–630. https://doi.org/10.1038/ejhg.2017.15

Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: an overview. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00019

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch--time interference in congenital amusia: evidence from an implicit test. *Attention, Perception & Psychophysics*, *72*(3), 763–774. https://doi.org/10.3758/APP.72.3.763

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

Schirmer, A., Escoffier, N., Cheng, X., Feng, Y., & Penney, T. (2016). Detecting Temporal Change in Dynamic Sounds: On the Role of Stimulus Duration, Speed, and Emotion. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02055

Schirmer, A., Striano, T., & Friederici, A. (2005). *Sex differences in the preattentive processing of vocal emotional expressions*. https://core.ac.uk/display/45977046

Stewart, L. (2011). Characterizing congenital amusia. *Quarterly Journal of Experimental Psychology (2006)*, *64*(4), 625–638. https://doi.org/10.1080/17470218.2011.552730

Tang, W., Wang, X.-J., Li, J.-Q., Liu, C., Dong, Q., & Nan, Y. (2018). Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research*, *363*, 62–69. https://doi.org/10.1016/j.heares.2018.03.004

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handbook of Clinical Neurology*, *129*, 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: evidence from a gating paradigm. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *59*, 84–94. https://doi.org/10.1016/j.cortex.2014.07.012

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital Amusia (or Tone-Deafness) Interferes with Pitch Processing in Tone Languages. *Frontiers in Psychology*, *2*, 120. https://doi.org/10.3389/fpsyg.2011.00120

Tillmann, B., Gosselin, N., Bigand, E., & Peretz, I. (2012). Priming paradigm reveals harmonic structure processing in congenital amusia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(8), 1073–1078. https://doi.org/10.1016/j.cortex.2012.01.001

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia*, *85*, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640, Part B*, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

Tillmann, B., Peretz, I., Bigand, E., & Gosselin, N. (2007). Harmonic priming in an amusic patient: The power of implicit tasks. *Cognitive Neuropsychology*, *24*(6), 603–622. https://doi.org/10.1080/02643290701609527

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., & Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *The Journal of the Acoustical Society of America*, *130*(6), 4089–4096. https://doi.org/10.1121/1.3658447

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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., Kesteren, E.-J. van, Doorn, J. van, Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1–19. https://doi.org/10.3758/s13423-017-1323-7

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., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory (Hove, England)*, *18*(6), 657–669. https://doi.org/10.1080/09658211.2010.501339

Zendel, B. R., Lagrois, M.-É., Robitaille, N., & Peretz, I. (2015). Attending to Pitch Information Inhibits Processing of Pitch Information: The Curious Case of Amusia. *The Journal of Neuroscience*, *35*(9), 3815–3824. https://doi.org/10.1523/JNEUROSCI.3766-14.2015

Zhang, C., & Shao, J. (2018). Normal pre-attentive and impaired attentive processing of lexical tones in Cantonese-speaking congenital amusics. *Scientific Reports*, *8*(1), 8420. https://doi.org/10.1038/s41598-018-26368-7

Zhang, C., Shao, J., & Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PloS One*, *12*(8), e0183151. https://doi.org/10.1371/journal.pone.0183151

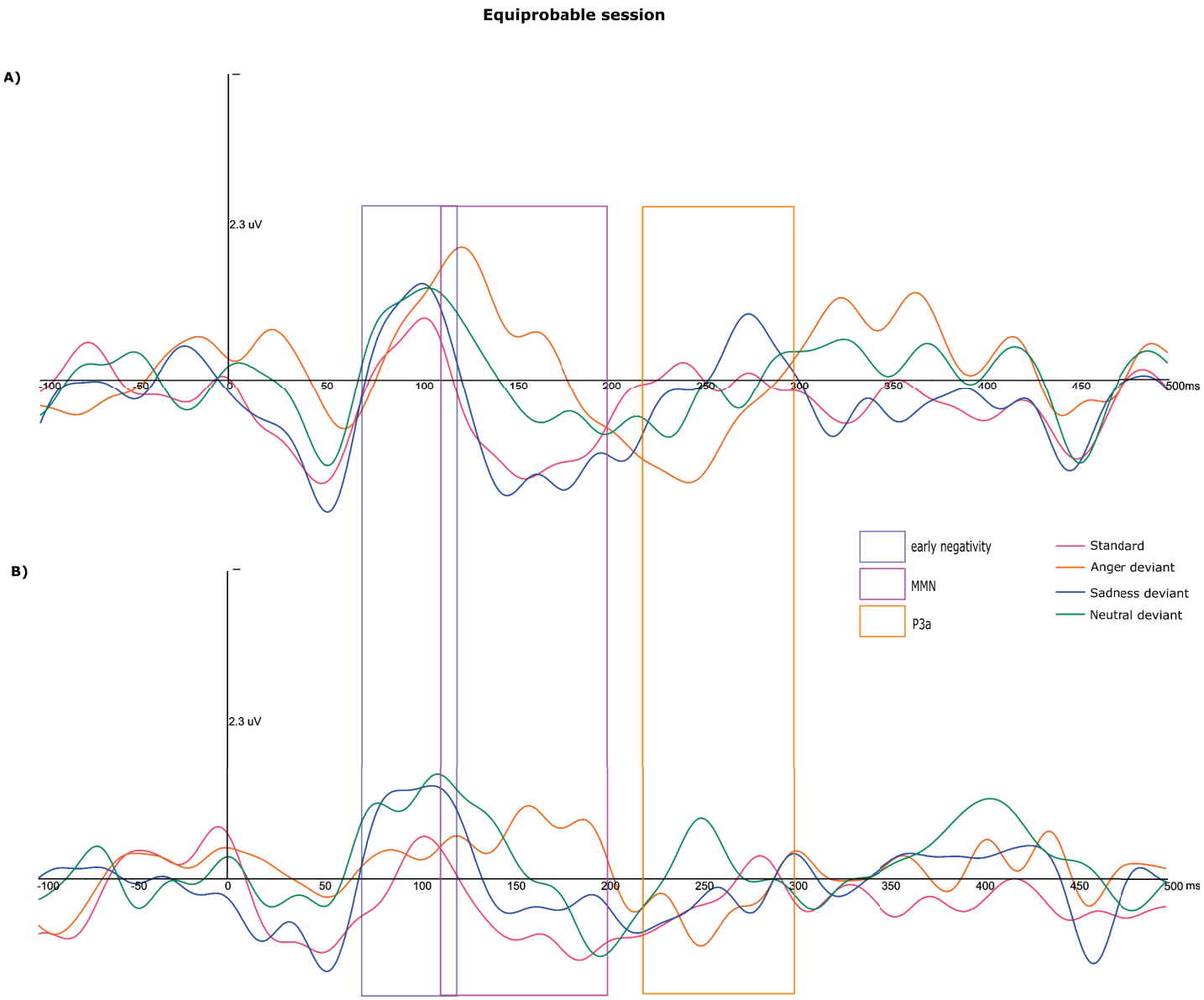

Figure S1: Average curve of ten fronto-central electrodes (Fp1, Fp2, F3, Fz, F4, FC1, FC2, C3, Cz, C4) of the response in the equiprobable session to the sounds that were used as a neutral standard, anger, sadness and neutral deviants in the main blocks, for controls (A) and amusics (B). Negativity is up. Emergence windows of the principal analysis are reported. The anger sound (and to a lesser extent the neutral deviant sound) elicited a negativity compared to the other equiprobable sounds in the latency range of the MMN, suggesting that MMNs were elicited within this sequence.

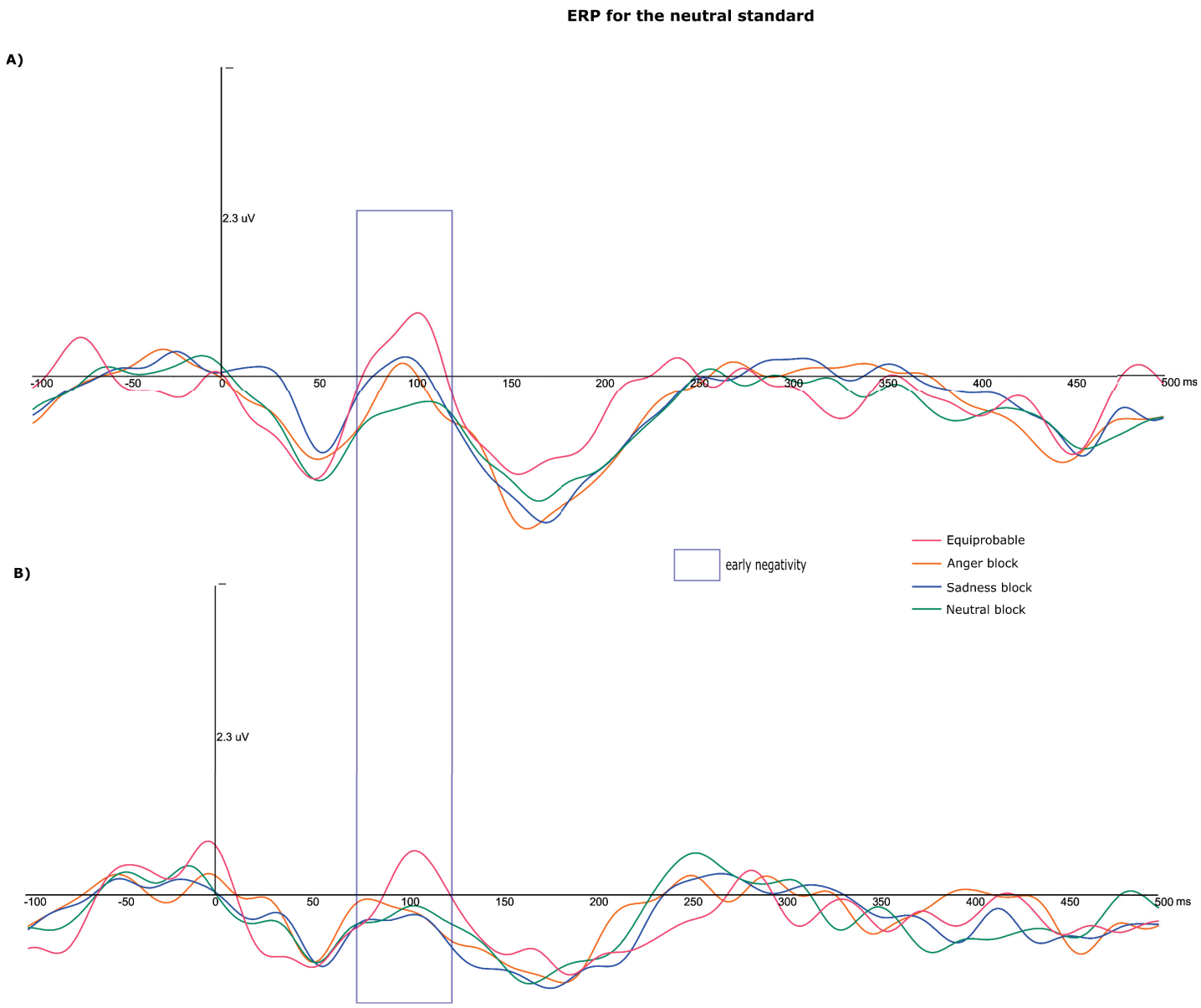

Figure S2: Average curve of ten fronto-central electrodes (Fp1, Fp2, F3, Fz, F4, FC1, FC2, C3, Cz, C4) of the response to a neutral standard in the four conditions (equiprobable session, block 1 with anger deviant, block 2 with sadness deviant, block 3 with neutral deviant), for controls (A) and amusics (B). Negativity is up. Emergence window or the early negativity in the principal analysis is reported. An early negativity at the latency of N1 was present only in the equiprobable session, suggesting that a habituation of N1 was occurring in the oddball blocks.

Supplementary analysis on midline electrodes (Fz, Cz, Pz)

Response to a neutral deviant

1. ***Early negativity***

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=1.3e+5). This model was 1.9 times better than the model with the main effects of Localization and Group (BF10=6.7e+4), and 8.7 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=1.5+4). The model with the main effect of Group showed no evidence (BF10=0.45). This was confirmed by a decisive specific effect of Localization (BFinclusion=9.6e+4), and no other specific effects (BFinclusion<0.43). The Group effect emerging in the second best model showed that amusics tended to have a smaller early negativity than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001).

2. ***MMN***

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=1.7e+11). This model was 1.6 times better than the model with the main effects of Localization and Group (BF10=1.04e+11), and 1.7 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=1+11). The model with the main effect of Group showed no evidence (BF10=0.51). This was confirmed by a decisive specific effect of Localization (BFinclusion=1.7e+11), and no other specific effects (BFinclusion<1.4). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001), amplitude at Cz was smaller than amplitude at Fz (pcorr<0.001).

3. *P3a*

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=4.1e+8). This model was 2.4 times better than the model with the main effects of Localization and Group (BF10=1.7e+8), and 15.8 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=2.6e+7). The model with the main effect of Group showed no evidence (BF10=0.36). This was confirmed by a decisive specific effect of Localization (BFinclusion=2.9e+8), and no other specific effects (BFinclusion<0.3). The Group effect emerging in the second best model showed that amusics tended to have a smaller P3a than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001), amplitude at Cz was smaller than amplitude at Fz (pcorr<0.001).

Response to a sadness deviant

1. *Early negativity*

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=2.1+6). This model was 1.5 times better than the model with the main effects of Localization and Group (BF10=1.4e+6), and 4.1 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=5.1e+5). The model with the main effect of Group showed no evidence (BF10=0.62). This was confirmed by a decisive specific effect of Localization (BFinclusion=1.7e+6), and no other specific effects (BFinclusion<0.6). The Group effect emerging in the second best model showed that amusics tended to have a smaller early negativity than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001).

2. *MMN*

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=3.9e+9). This model was 2.1 times better than the model with the main effects of Localization and Group (BF10=1.9e+9), and 9.7 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=4.02+8). The model with the main effect of Group showed no evidence (BF10=0.43). This was confirmed by a decisive specific effect of Localization (BFinclusion=2.9e+9), and no other specific effects (BFinclusion<0.4). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001), amplitude at Cz was smaller than amplitude at Fz (pcorr<0.001).

3. *P3a*

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=6.2e+6). This model was 1.7 times better than the model with the main effects of Localization and Group (BF10=3.6e+6), and 12.1 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=5.1e+5). The model with the main effect of Group showed no evidence (BF10=0.49). This was confirmed by a decisive specific effect of Localization (BFinclusion=4.6e+6), and no other specific effects (BFinclusion<0.4). The Group effect emerging in the second best model showed that amusics tended to have a smaller P3a than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001), amplitude at Cz was smaller than amplitude at Fz (pcorr=0.012).

Response to an anger deviant

1. MMN

After comparison to the null model, the best model showing decisive evidence was the model with the main effect of Localization (BF10=8.5e+10). This model was 1.7 times better than the model with the main effects of Localization and Group (BF10=4.9e+10), and 11.6 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=7.3+9). The model with the main effect of Group showed no evidence (BF10=0.48). This was confirmed by a decisive specific effect of Localization (BFinclusion=6.4e+10), and no other specific effects (BFinclusion<0.4). The Group effect emerging in the second best model showed that amusics tended to have a smaller MMN than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001), amplitude at Cz was smaller than amplitude at Fz (pcorr=0.016).

2. P3a

After comparison to the null model, the best model showing decisive evidence was the model with the main effects of Localization and Group (BF10=1.8e+10). This model was 1.2 times better than the model with the main effect of Localization (BF10=1.5e+10), and 7.8 times better than the model with the main effects of Localization and Group and the interaction between the two (BF10=2.3e+9). The model with the main effect of Group showed no evidence (BF10=1.3). This was confirmed by a decisive specific effect of Localization (BFinclusion=1.02e+10), and no other specific effects (BFinclusion<0.9). The Group effect emerging in the second best model showed that amusics had a smaller P3a than controls. According to post-hoc tests, amplitude at Pz was smaller than amplitudes at Fz and Cz (both pcorr<0.001).

V. Musical emotions perception following a brain lesion: impact of lesion side

A. Objectives

Previous studies have investigated brain-damaged patients to better understand brain functioning and the potential role of brain regions (Stewart et al., 2006). In particular, multiple studies have investigated emotion perception in brain-damaged patients controls (Borod, Bloom, Brickman, Nakhutina, & Curko, 2002; Charbonneau, Scherzer, Aspirot, & Cohen, 2003; Cheung, Lee, Yip, King, & Li, 2006; Harciarek, Heilman, & Jodzio, 2006). These studies proposed two hypotheses regarding emotion perception in brain-damaged patients: the right hemisphere hypothesis suggests that the right hemisphere is dominant for emotion processing in general, while the valence hypothesis suggests that positive emotions could be processed by the left hemisphere whereas negative emotions would be processed by the right hemisphere. However, these two hypotheses were mostly derived from studies with facial and vocal material but only few studies have investigated musical emotion perception in brain-damaged patients.

In this framework, Study 3 investigated musical emotion perception in unilateral brain-damaged patients. As in Study 1 and Lévêque et al. (2018), we used a two-step paradigm with explicit emotion recognition and implicit intensity ratings. We tested left and right brain-damaged patients as well as healthy comparison participants. To determine which of the two previously suggested hypotheses is validated with musical emotions, we tested patients with various brain lesion locations for conclusions over a large set of brain regions. To tease apart general emotion perception deficits from specific musical deficits, we also used a facial emotion perception task.

B. Study 3: Recognition of musical emotions and their perceived intensity after unilateral brain damage

CORTEX 130 (2020) 78–93

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special Issue "Brain and cognitive asymmetry in clinical disorders": Research Report

Recognition of musical emotions and their perceived intensity after unilateral brain damage

Agathe Pralus [a,b,1,*], *Amy Belfi* [d,1], *Catherine Hirel* [a,b,e,1], *Yohana Lévêque* [a,b], *Lesly Fornoni* [a,b], *Emmanuel Bigand* [f], *Julien Jung* [a,b,e], *Daniel Tranel* [c], *Norbert Nighoghossian* [b,e,g], *Barbara Tillmann* [a,b,1] *and Anne Caclin* [a,b,1]

[a] *Lyon Neuroscience Research Center; CNRS, UMR5292; INSERM, U1028; Lyon, France*
[b] *University Lyon 1, Lyon, France*
[c] *Department of Neurology, University of Iowa, Iowa City, IA, USA*
[d] *Department of Psychological Science, Missouri University of Science and Technology, Rolla, MO, USA*
[e] *Hôpital Neurologique Pierre Wertheimer, Hospices Civils de Lyon, Bron, France*
[f] *LEAD, CNRS, UMR 5022, University of Bourgogne, Dijon, France*
[g] *CREATIS, CNRS, UMR5220, INSERM, U1044, University Lyon 1, France*

ARTICLE INFO

Article history:
Received 30 December 2019
Reviewed 21 April 2020
Revised 27 May 2020
Accepted 29 May 2020
Published online 12 June 2020

Keywords:
Brain lesion
Emotion perception
Music
Right hemisphere hypothesis
Valence hypothesis

ABSTRACT

For the hemispheric laterality of emotion processing in the brain, two competing hypotheses are currently still debated. The first hypothesis suggests a greater involvement of the right hemisphere in emotion perception whereas the second hypothesis suggests different involvements of each hemisphere as a function of the valence of the emotion. These hypotheses are based on findings for facial and prosodic emotion perception. Investigating emotion perception for other stimuli, such as music, should provide further insight and potentially help to disentangle between these two hypotheses. The present study investigated musical emotion perception in patients with unilateral right brain damage (RBD, n = 16) or left brain damage (LBD, n = 16), as well as in matched healthy comparison participants (n = 28). The experimental task required explicit recognition of musical emotions as well as ratings on the perceived intensity of the emotion. Compared to matched comparison participants, musical emotion recognition was impaired only in LBD participants, suggesting a potential specificity of the left hemisphere for explicit emotion recognition in musical material. In contrast, intensity ratings of musical emotions revealed that RBD patients underestimated the intensity of negative emotions compared to positive emotions, while LBD patients and comparisons did not show this pattern. To control for a potential generalized emotion deficit for other types of stimuli, we also tested facial

Abbreviations: RBD, Right-brain damage; LBD, Left-brain damage; MBEA, Montreal Battery of Evaluation of Amusia; PDT, Pitch Discrimination Threshold; MMSE, Mini Mental State Examination; WAIS-IV, Wechsler Adult Intelligence Scale; WCST, Wisconsin Card Sorting Test; MEC, Montreal Evaluation of Communication; AVLT, Auditory-Verbal Learning Test; BNT, Boston Naming Test; COWA, Controlled Oral Word Association.

* *Corresponding author*. CRNL, 95 Bd Pinel, 69675, Bron Cedex, France.
E-mail address: agathe.pralus@inserm.fr (A. Pralus).

[1] Equally contributing authors.

https://doi.org/10.1016/j.cortex.2020.05.015
0010-9452/© 2020 Elsevier Ltd. All rights reserved.

emotion recognition in the same patients and their matched healthy comparisons. This revealed that emotion recognition after brain damage might depend on the stimulus category or modality used. These results are in line with the hypothesis of a deficit of emotion perception depending on lesion laterality and valence in brain-damaged participants. The present findings provide critical information to disentangle the currently debated competing hypotheses and thus allow for a better characterization of the involvement of each hemisphere for explicit emotion recognition and their perceived intensity.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Studying perception in brain-damaged patients has contributed to a better understanding of various brain functions, including emotion perception. Seminal studies have investigated brain-damaged patients' emotion processing in faces and voices, aiming for a better understanding of patients' communication with their social environment (Borod, 1992; Peretz, 1990; Peretz, Gagnon, & Bouchard, 1998; Sackeim et al., 1982). For facial emotion recognition, results consistently describe increased difficulties for brain-damaged patients to recognize facial emotions in comparison with controls (Borod, Bloom, Brickman, Nakhutina, & Curko, 2002; Charbonneau, Scherzer, Aspirot, & Cohen, 2003; Cheung, Lee, Yip, King, & Li, 2006; Harciarek, Heilman, & Jodzio, 2006). However, the degree of impairment seems to depend on lesion location (Yuvaraj, Murugappan, Norlinah, Sundaraj, & Khairiyah, 2013). For example, lesions in subcortical structures, such as thalamus and basal ganglia, have been associated only with a small decrease in facial emotion recognition compared to controls (Cheung et al., 2006). When the lesion involves cortical regions, the degree of impairment for facial emotion recognition varies according to the damaged area, with more impairment for anterior brain lesions than posterior lesions, in particular for negative valence emotions (Harciarek & Heilman, 2009).

Numerous studies have investigated the effect of lesion laterality on (facial and prosody) emotion recognition, with findings resulting in two contrasting hypotheses. The *Right Hemisphere Hypothesis* suggests that the right hemisphere is dominant for emotion processing independently of the type of emotion, while the *Valence Hypothesis* suggests that positive emotions are preferentially processed in the left hemisphere whereas negative emotions are preferentially processed in the right hemisphere (Abbott, Cumming, Fidler, & Lindell, 2013; Adolphs, Jansari, et al., 2001). In support of the Right Hemisphere Hypothesis, numerous studies have reported stronger emotion recognition impairments in right brain-damaged (RBD) patients compared to left brain-damaged (LBD) patients and to controls (Adolphs, Tranel, & Damasio, 2001; Borod, Bloom, Brickman, Nakhutina, & Curko, 2002; Charbonneau et al., 2003; Harciarek et al., 2006; Kucharska-Pietura, Phillips, Gernand, & David, 2003a; Tippett et al., 2018), without potential interaction with the valence of the emotions. Notably, a meta-analysis on facial emotion recognition in brain-damaged patients suggests more involvement of the right hemisphere for emotion perception as RBD patients were more impaired than were LBD patients. However, it also suggests a right lateralization specific for negative valence emotion perception, but no lateralization for positive valence emotion perception (Abbott et al., 2013). In support of the Valence hypothesis, recent studies reported a specific deficit of RBD patients for negative emotions (Braun, Traue, Frisch, Deighton, & Kessler, 2005; Nijboer & Jellema, 2012). This finding is in line with the observation that right-hemisphere lesions were associated to pathological laughing and euphoric mood change, while left-hemisphere lesions were associated to pathological crying (Sackeim et al., 1982). Yet other studies also reported impaired emotion perception for both RBD and LBD patients without lateralization of the deficit and no clear link with the valence of the emotion (Abbott, Wijeratne, Hughes, Perre, & Lindell, 2014; Braun, Traue, Frisch, Deighton, & Kessler, 2005; Cheung et al., 2006). Overall, the results regarding brain lateralization and emotion perception are still unclear, with no clear-cut evidence for one specific hypothesis, at least when emotion processing was studied with face stimuli (Abbott et al., 2013).

As emotions can be communicated not only via visual cues, but also auditory cues, some studies have investigated vocal emotion perception, such as emotional prosody, in brain-damaged patients (see Yuvaraj et al., 2013 for a review). Most studies using language and vocalization materials were in support of the Right Hemisphere Hypothesis, reporting greater impairment of RBD patients for emotional prosody recognition than LBD patients and controls (Borod et al., 2002; Charbonneau et al., 2003; Harciarek et al., 2006; Kucharska-Pietura, Phillips, Gernand, & David, 2003). Another study confirmed the deficit for RBD patients, which was larger than for LBD patients, but for this one study, the LBD patients also had a mild impairment for emotional prosody compared to controls (Kucharska-Pietura et al., 2003).

Overall, numerous studies investigating emotion perception in brain-damaged patients with facial and prosody materials have revealed complex patterns of impairments depending both on lesion lateralization and lesion localization. Only few studies have investigated musical emotions in brain-damaged patients, even though emotions are an important motivation for music listening (Egermann, Fernando, Chuen, & McAdams, 2014). In the musical domain, it has been shown that emotion recognition can be

preserved in patients even when their musical structure perception is impaired (Peretz et al., 1998). Several case reports have revealed that brain damage can result in musical anhedonia, a specific loss of experience of pleasure for music whereas emotion recognition is intact (Belfi, Evans, Heskje, Bruss, & Tranel, 2017; Griffiths, Warren, Dean, & Howard, 2004; Satoh, Nakase, Nagata, & Tomimoto, 2011, 2016). For instance, a patient with a right inferior parietal lobe infarct did not perceive any emotion when listening to music, but its music perception and emotion recognition were preserved (Satoh et al., 2011). Case reports (Gosselin, Peretz, Johnsen, & Adolphs, 2007; Griffiths et al., 2004; Satoh et al., 2016) and group studies (Gosselin, Peretz, Hasboun, Baulac, & Samson, 2011; Jafari, Esmaili, Delbari, Mehrpour, & Mohajerani, 2017; Khalfa et al., 2007, 2008) have also reported deficits of music emotion recognition in brain-damaged patients. For instance, patients with unilateral medial temporal lesions showed more difficulties to recognize musical emotions, especially for fearful stimuli (Gosselin et al., 2011), with no clear association to one side of the lesion. Regarding the potentially differentiated roles of the two hemispheres for musical emotion recognition, it has been observed that patients with right temporal lobe lesions have more difficulties in recognizing emotions in music than patients with left temporal lobe lesions (Jafari et al., 2017). More precisely, these RBD patients had greater difficulties in recognizing negative emotions such as sadness compared to LBD patients (Jafari et al., 2017; Khalfa et al., 2007), whereas LBD patients had greater difficulties in recognizing positive emotions such as happiness (Khalfa et al., 2007), in keeping with the predictions of the Valence Hypothesis. These results were consistent with the studies of musical emotion perception in healthy participants that demonstrate a lateralization of this perception according to the valence of the emotion (Altenmüller, Schürmann, Lim, & Parlitz 2002; Tsang, Trainor, Santesso, Tasker, & Schmidt, 2001). In another study, RBD patients overestimated the arousal for happiness in music, compared to LBD patients (Khalfa et al., 2008). However, when asked to judge emotional dissimilarities in musical excerpts in terms of arousal and valence instead of emotion recognition, patients with left or right unilateral medial temporal lesions did not show any deficit (Dellacherie, Bigand, Molin, Baulac, & Samson, 2011). This was consistent with results reported for healthy participants that showed no clear pattern regarding brain lateralization of valence for musical perception (Khalfa, Schon, Anton, & Liégeois-Chauvel, 2005).

Variability in the results regarding musical emotions processing of previous studies might be related to the diversity of experimental paradigms. Some studies used musical emotion categorization tasks (Gosselin et al., 2011; Jafari et al., 2017; Peretz et al., 1998) and others required ratings of the emotions' intensity (Gosselin et al., 2007; Griffiths et al., 2004) or ratings of valence and arousal (Dellacherie et al., 2011; Gosselin et al., 2007; Khalfa et al., 2008; Satoh et al., 2011). As previously suggested in studies with facial and prosodic material in healthy participants and unilateral brain-damaged patients (Abbott et al., 2013; Borod et al., 2002; Demaree, Everhart, Youngstrom, & Harrison, 2005), these task effects could reflect the distinction between the recognition and the actual experience of emotions. Indeed, some studies have suggested that the right hemisphere hypothesis would be more strongly associated with emotion recognition, i.e., with a cognitive or intentional process, whereas the valence hypothesis would be more strongly associated with automatic processing of the emotion and thus being closer to the emotional experience of participants (Abbott et al., 2013; Borod et al., 2002; Demaree et al., 2005). These two modes of emotional processing would rely on different anatomical substrates, with emotion recognition associated with hemispheric asymmetries in posterior and temporal regions, whereas emotion experience would be associated with hemispheric asymmetries in more frontal regions (Abbott et al., 2013; Borod, 1992). To further contribute to the distinction between these two processes in association with the two hypotheses of emotion processing in music, tn the present study, a two-task paradigm was used to assess both musical emotion categorization and intensity ratings of these emotions in unilateral brain-damaged patients and matched healthy comparison participants. In this paradigm, participants were required to choose the recognized emotion among four possibilities (Joy, Fear/Anger, Sadness, or Neutrality/Serenity) in musical, and then to rate the intensity of this emotion on a five-point scale. The intensity of emotions can be done without verbal or categorical representation of the emotion as a global appreciation of the stimulus or a fuzzy representation of emotion suffices (Lévêque et al., 2018). Intensity ratings of emotions reflect a more implicit perception of the emotion and could be closely related to the actual feeling of this emotion (Hirel et al., 2014). In previous studies, this paradigm has allowed us to show that emotion recognition, but not intensity ratings of these emotions, is disrupted in congenital amusia, both for musical material (Lévêque et al., 2018) and emotional prosody material (Pralus et al., 2019). This paradigm also allowed for the identification of emotion recognition deficits without intensity ratings deficits in single cases of brain-damaged patients: in one patient with musical material (Hirel et al., 2014) and in another patient with emotional prosody material (Bourgeois–Vionnet, Moulin, Hermier, Pralus, & Nighoghossian, 2020). Furthermore, participants were tested with the same paradigm but using face stimuli, to assess the specificity of the reported effects to the musical domain. Indeed, brain imaging studies in healthy participants have demonstrated shared brain networks for emotion perception with music, vocalization and also face material, especially for fear recognition (Aubé, Angulo-Perkins, Peretz, Concha, & Armony, 2015; Koelsch et al., 2013; Paquette et al., 2018). Thus impairments of facial and vocal emotion perception might cooccur with impaired musical emotion perception.

The aim of the present study was to investigate musical emotion recognition and its perceived intensity after unilateral brain damage, in particular to determine the potential effect of lesion side on emotion perception. In contrast to previous group studies investigating musical emotions in brain-damaged patients, we did not restrict the patient selection to lesions encompassing mesial temporal structures (including amygdala or parahippocampus) (Dellacherie et al., 2011; Frühholz, Trost, & Grandjean, 2014; Gosselin et al., 2006, 2011), but included patients with a large variability of lesion locations aiming for a wider conclusion about the link between lesion side and emotion perception. We compared

the recognition of musical emotions and their rated intensity in RBD patients, LBD patients, and healthy comparison participants. To tease apart general emotion recognition deficits from specific auditory or musical deficits, we also used a facial emotion recognition task with its subsequent intensity ratings. Music perception abilities were also assessed to analyse their potential contribution and/or dissociation to musical emotion perception.

2. Materials and methods

2.1. Participants

Thirty-two brain-damaged patients and 28 healthy comparison participants were included in the study (Table 1). Thirty-four participants were recruited in Lyon and its surroundings (France), and 26 participants in Iowa (USA). Inclusion criteria for patients were the presence of a focal unilateral brain damage involving the cortex, without prior psychiatric disease, severe cognitive disorder, severe hearing or visual loss. All patients were tested in the chronic phase of their condition.

(more than 3 months after lesion onset). In total, 16 left brain-damaged patients (eleven from France, five from Iowa) and 16 right brain-damaged patients (ten from France, six from Iowa) were included. The 21 French patients were recruited among the patients of the stroke unit of the neurological hospital in Lyon, France. They presented a unilateral ischemic stroke in the right or left middle cerebral artery territory, confirmed by MRI. The 11 patients from Iowa (USA) presented focal brain damage due to vascular lesions (n = 7, including four in the territory of the middle cerebral artery, one in the anterior cerebral artery, one in the internal carotid artery, and one in the vertebral artery), surgical resection of a frontal tumor (n = 1), temporal lobectomy for epilepsy relief (n = 3). They all (except one) underwent a high resolution MRI to localize their lesions (see Table 2). Thirteen healthy comparisons were recruited in France, and fifteen healthy comparisons were recruited in the USA. They were matched to patients for age, gender, education level, and music training. Study procedures were approved by the appropriate ethics committee on both sites and participants were paid for their participation. All participants' consent was obtained according to the Declaration of Helsinki.

2.2. Neuropsychological assessment

Prior to the main experiment, all participants were tested with an audiometry, the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003) to diagnose amusia, and a Pitch Discrimination Threshold (PDT) test (Tillmann et al., 2009). A participant was considered as amusic if he/she had a global MBEA score below 22.4/30 for participants under 60 years and 21.6/30 for participants aged over sixty years (see Table 1) (http://www.brams.umontreal.ca/plab/publications/article/57#extras) (Peretz et al., 2003).

To assess general cognitive abilities of patients, neuropsychological measures were collected before the testing session (Tables S1 & S2). Different, though globally equivalent neuropsychological tests were used in the two recruitment sites (France and Iowa). To test general cognitive functioning, French patients underwent the Mini Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975), and the American patients underwent the WAIS-IV (Wechsler Adult Intelligence Scale) (Hartman, 2009) for full-scale IQ (we also report sub-scores for working memory, WMI, and processing speed, PSI), the WCST (Wisconsin Card Sorting Task, Nelson, 1976) for perseverative errors (PE) and categories completed (CAT). To test verbal abilities, the French patients underwent lexical and categorical verbal fluencies, and the Montreal Evaluation of Communication (MEC) for the comprehension of linguistic prosody and emotional prosody, and the American patients were tested with an auditory-verbal learning test (AVLT), the Boston Naming Test (BNT, Kaplan, Goodglass, Weintraub, & Goodglass, 1983), and the Controlled Oral Word Association (COWA, Loonstra, Tarlow, & Sellers, 2001). The French patients also were administrated a depression scale test (Hamilton, 1960).

2.3. Stimuli

Forty musical excerpts were selected from the Western classical repertoire (Table S3). All excerpts were orchestrated instrumental stimuli, without voice, lasted 20 s, and were aimed to be representative of four emotions in real recordings (see Bigand, Filipic, & Lalitte, 2005; Filipic, Tillmann, & Bigand, 2010; Lévêque et al., 2018; Liégeois-Chauvel et al., 2014). In this selection of stimuli, ten excerpts related to joy (e.g., an excerpt from Beethoven's Piano, sonata 32, mvt 2), ten to sadness (e.g., an excerpt from Shostakovitch's Symphony 15, Adagio), ten to fear/anger (e.g., an excerpt from Chopin's Prelude, op.28, no.22), ten to serenity (e.g., an excerpt from Scarlatti's Sonata A for Harpsichord). Thus, there were two positive valence emotion categories and two negative valence emotion categories, with two high arousal emotion categories and two low arousal emotion categories.

For the visual task, forty photos of faces were selected from Ekman and Friesen (1976) (Lévêque et al., 2018). All photos were in black and white. They appeared on the screen for two seconds. To match the musical material, ten faces were related to joy, ten to sadness, ten to fear, and ten were emotionally neutral, as in Hirel et al. (2014) and Lévêque et al. (2018). Neutrality was used instead of serenity because serenity is difficult to recognize on a face.

2.4. Procedure

In each trial, participants listened to or watched a stimulus and were then asked to select the recognized emotion from four options (joy, serenity (music)/neutral (faces), sadness, fear/anger). During the tasks, only the word "fear" appeared on the screen. However, participants were informed at the beginning of the experiment that this category in the musical task corresponded to anger and fear. Indeed, anger and fear can be evoked by the same musical excerpts depending on perspective taken (see Hirel et al., 2014; Johnsen, Tranel,

Table 1 – Demographic data of participants. Standard deviations are indicated in parentheses. Group comparisons use ANOVAs with group (Comparisons, RBD patients, LBD patients) as between-participants factor, except for sex ratio and laterality where a Chi2 test was used. MBEA (Montreal Battery for the Evaluation of Amusia, Peretz, Champod, & Hyde, 2003) score = average score of the six subtests (scale, contour, interval, rhythm, meter, memory), significant difference between groups (in bold): LBD patients have significantly lower MBEA scores compared to healthy comparisons ($p = .022$) according to a Fisher-LSD post-hoc test. PDT: Pitch Discrimination Threshold (Tillmann, Schulze, & Foxton, 2009).

	HEALTHY COMPARISONS (N = 28)	RBD PATIENTS (N = 16)	LBD PATIENTS (N = 16)	P-VALUE (GROUP COMPARISON)
Sex ratio (M/F)	11/17	6/10	10/6	.26
Age (years)	58.3 (±9.9)	56 (±10.8)	67,8 (±11.7)	.21
*Laterality**	2L 24R	1L 15R	1L 15R	.98
Education (years)	14.5 (±3.4)	12.2 (±3.7)	12.9 (±3.5)	.15
*Musical education** (years)*	2 (±3.4)	.8 (±2.3)	5.6 (±14)	.24
Time since stroke (months)	NA	30.6 (±39.3)	67.4 (±78.6)	.19 (RBD vs. LBD)
Lesion size (mL)	NA	15.4 (±17.9)	21.2 (±18.4)	.55 (RBD vs. LBD)
MBEA score (max. score = 30)	25.2 (±2.1)	24 (±2.8)	23.3 (±3)	**.04**
*PDT*** (semi-tones)*	.74 (±.92)	1.31 (±1.25)	1.78 (±2.6)	.13

*Missing data for 2 healthy comparisons.
**Missing data for 5 healthy comparisons and 5 patients.
***Missing data for 6 patients.

Lutgendorf, & Adolphs, 2009; Lévêque et al., 2018). After having given their response, they were asked to rate the intensity of the emotion evoked by the musical excerpt or the face from 1 (not intense) to 5 (very intense), except for face stimuli judged as neutral. After the intensity rating response, the following stimulus was automatically played after a variable delay of 2500 ms on average (ranging from 2000 to 3000 ms). The stimuli were presented in two blocks: music in one and faces in another. The presentation order of the two blocks was counterbalanced across participants. The participant was allowed taking a small break between the two blocks. Within a block, the presentation order of the stimuli was randomized for each participant, with the constraint that a given emotion cannot be presented more than three times in a row. For both blocks (music and faces), participants were not asked to distinguish between felt and perceived emotion. Indeed, it was shown that this distinction can be complex to perform (Niedenthal, 2007; Scherer, 2004). Presentation software (Neurobehavioral systems, Albany, CA, USA) was used to present the stimuli to the participants and to record responses on a keyboard. The duration of the experiment was 20 min.

2.5. *Data analyses*

For each participant and emotion, separately for musical excerpts and faces, the percentages of correct responses (categorization score) and the average ratings of intensity for correctly categorized trials were calculated. Each dependent variable was analyzed with a 3 × 4 ANOVA with Group (LBD patients vs. RBD patients vs. comparison participants) as the between-participants factor and Emotion (Joy, Sadness, Fear, Serenity/Neutral) as the within-participant factor. For intensity ratings for facial emotions, the factor Emotion had only three levels (Joy, Sadness, and Fear), as intensity ratings were not performed for neutral stimuli. The Greenhouse-Geisser correction was applied if appropriate and corrected degrees of freedom are reported. We calculated Pearson-correlation between categorization scores and MBEA scores within each participant group (RBD, LBD, comparisons) and over the three groups. Similarly, we calculated Pearson-correlation between categorization scores and the PDT within each participant group (RBD, LBD, comparisons) and over the three groups, even though PDT data were missing for six participants (3 RBD and 3 LBD patients).

We run an additional ANOVAs for music material on categorization scores and intensity ratings with MBEA score as a covariate, to further investigate a possible link between musical perception and memory abilities (as measured in the MBEA) and emotional processing in the three groups of participants.

As the music material had been constructed in France, we also tested for potential cross-cultural differences between participants by analyzing the data of comparison participants with a 2 × 4 ANOVA with Site (France vs. USA) as a between-participants factor and Emotion (Joy, Sadness, Fear, Serenity/Neutral) as the within-participant factor, for recognition performance and intensity ratings of music and for recognition performance of face material respectively. For intensity ratings of face material, a 2 × 3 ANOVA was performed as the factor Emotion did not include Neutrality.

To test for potential effects of slightly different patient recruitment criteria on the two sites (only middle cerebral artery stroke patients were recruited in France, whereas patients with more diverse lesion etiologies were recruited in the USA), we analyzed the patient data of each dependent variable with a 2 × 2 × 4 (or 2 × 2 × 3) ANOVA with Site (France vs. USA) and Lesion Laterality (RBD vs. LBD) as between-participants factors, and Emotion (Joy, Sadness, Fear, and Serenity/Neutral where appropriate) as the within-participant factor.

For all analyses, post-hoc analyses for significant effects or interactions were carried out using Fisher LSD tests.

Individual patient data for musical excerpts and faces (percentages of correct responses and average intensity ratings for

Table 2 – Individual data on lesion localization, correct categorizations and intensity ratings of music and faces, and MBEA scores. F1–F21: French patients (F11 is described in detail in Hirel et al., 2014), A1-A11: American patients. T = temporal, F = frontal, P = parietal, I = insula, O = occipital, BG = basal ganglia. % Corr: mean correct categorization. Int.: mean intensity ratings. Data below or above the cutoffs are in bold.

PATIENTS	LESION SIDE	LESION LOCALISATION						MUSIC		FACES		MEAN MBEA (MAX=30)
		T	F	P	I	O	BG	% CORR.	INT.	% CORR.	INT	
F1	Right				x		x	62.5	3.27	**70**	**1.90**	22.5
F2	Right	x		x	x			77.5	3.82	90	4.06	24.17
F3	Right	x		x	x			75	4.18	85	3.43	27
F4	Right	x	x		x			**55**	3.67	85	**4.47**	24.17
F5	Right			x	x			77.5	3.55	90	4.13	24.5
F6	Right	x						80	3.33	90	3.07	27.5
F7	Right	x	x		x			67.5	3.31	87.5	3.40	26.83
F8	Right			x				82.5	4	**70**	3.21	23.5
F9	Right		x	x				75	4.42	80	3.33	22.67
F10	Right			x	x			57.5	3.43	77.5	**2.40**	23.33
F11	Right	x						77.5	**1.48**	80	3.3	**21.5**
A1	Right	x						82.5	2.42	95	3.89	**19.17**
A2	Right						x	72.5	3.08	85	3.7	23
A3	Right						x	65	2.52	**67.5**	3.05	**21**
A4	Right		x					92.5	3.37	97.5	3.47	28.67
A5	Right	Missing data						60	3.73	95	3.63	**21.33**
F12	Left	x		x	x			**52.5**	2.42	95	3.43	**18.5**
F13	Left	x		x	x			85	3.48	92.5	3.90	24.17
F14	Left		x		x			65	3.11	82.5	3.45	26.67
F15	Left	x		x	x			72.5	3.44	85	3.86	22.83
F16	Left		x		x			72.5	3.33	82.5	3.33	25.33
F17	Left			x	x			**42.5**	3.56	82.5	3.36	22
F18	Left		x	x	x			**50**	4.34	80	**4.49**	**17.67**
F19	Left		x		x			62.5	**4.96**	80	3.94	**18.67**
F20	Left	x						67.5	3.17	87.5	3.50	26.17
F21	Left				x			85	3.69	82.5	3.86	26.17
A6	Left	x						**50**	3.56	92.5	3.91	**21.17**
A7	Left		x		x			82.5	2.87	90	3.70	24
A8	Left			x				72.5	4.29	87.5	3.93	25.5
A9	Left					x		70	3	95	2.92	22.5
A10	Left	x						**55**	3.49	85	**1.90**	26
A11	Left					x		77.5	3.42	87.5	2.51	26.67
TOTAL	Right	7	4	6	7	0	3					
TOTAL	Left	6	5	6	10	2	0					
COMPARISONS MEAN								78.92	3.56	88.39	3.44	
HIGH CUTOFF									4.86		4.38	
LOW CUTOFF								55.53	2.26	74.04	2.50	22.4 (<60 years) 21.6 (>60 years)

correctly categorized trials) were also analyzed. For percentages of correct responses, individual data were compared to a cutoff score corresponding to the comparisons' mean minus two standard deviations. For average intensity ratings, individual data were compared to a cutoff score corresponding to the comparisons' mean minus two standard deviations (low cutoff) and to a cutoff score corresponding to the comparisons' mean plus two standard deviations (high cutoff).

2.6. *Transparency and openness promotion of the study*

We report in the 'Participants' section how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

We were not the owners of most of the stimuli we used in the study. Readers seeking access to the stimuli may contact A. Pralus to obtain the contacts of the research teams who originally created the stimuli and collaborated with us either in the present study or a previous one or have made public access already. Material and stimuli of the MBEA (Peretz et al., 2003) are accessible on https://www.peretzlab.ca/knowledge_transfer/. The Pitch Discrimination test has been developed by Jessica Foxton and used in the following publications or our team: Hirel et al., 2014; Lévêque et al., 2018; Pralus et al., 2019; Tillmann et al., 2009. Stimuli from the MMSE test are hold by the Folstein group (Folstein et al., 1975). Instructions for the WAIS examination are commercialized by the Wechsler group (Hartman, 2009). Stimuli from the MEC protocol are commercialized by Ortho Edition (Joanette et al., 2004. Protocole Montréal d'évaluation de la communication (MEC). Isbergues, France: Ortho-Edition). The stimuli of the Boston Naming Test are commercialized by Pearson Clinical (Kaplan et al., 1983. Boston Naming Test-Second Edition). The stimuli of Controlled Oral Word Association are owned by the Psychological

Assessment Resources (Loonstra et al., 2001. PAR, https://www4.parinc.com/Products/PermissionsAndLicensing.aspx). The depression scale test is available as an appendix in the original research article (Hamilton, 1960). Musical stimuli for the emotion categorization test have been selected by Emmanuel Bigand and Philippe Lalitte (University of Burgundy, LEAD–CNRS 5022, Dijon, France), and used in the following collaborative publication: Leveque et al. (2018). Copyright for face stimuli (Ekman & Friesen, 1976) is held by Paul Ekman Group.

The conditions of our ethics approval do not permit public archiving of anonymized study data. Readers seeking access to the data should contact A. Pralus. Access will be granted to named individuals in accordance with ethical procedures governing the reuse of clinical data, including completion of a formal data sharing agreement and approval of the local ethics committee.

No part of the study procedures was pre-registered prior to the research being conducted.

3. Results

3.1. Neuropsychological data

Tables S1 and S2 show results of the neuropsychological assessment for patients from both recruitment sites. For general cognitive functioning, only three French LBD patients were slightly cognitively impaired (MMSE scores between 23 and 24), no American patient had an impairment (all WAIS scores between 70 and 130), except one American RBD patient who was below the norms of the WCST-PE (but not impaired for the WCST-CAT). For verbal abilities in French patients, two RBD and one LBD patients had a deficit for lexical fluencies (scores lower than 8.09, age-adjusted cutoff) and one RBD patient had a deficit for categorical fluencies (score lower than 20.46, age-adjusted cutoff), no patient was below the norm for emotional and linguistic prosody (MEC battery). For verbal abilities in American patients, only one RBD patient had a deficit for AVLT (score lower than 6.8), no patient had a deficit for BNT and COWA. For the depression scale (only French patients were tested), 6 RBD and 7 LBD patients had scores below the norm (scores lower than 7).

Overall, these neuropsychological tests revealed that the patients included in the study were not severely cognitively impaired, and potential deficits observed in our paradigm would most likely not be due to a more general deficit of cognition. The depression scale revealed that some patients were not in the norm (6 RBD and 7 LBD patients), which is common in brain-damaged patients. However, most importantly, depression scores were similar in LBD and RBD patients, thus depression scores cannot explain potential group differences between the two patient groups in the other tasks.

Regarding music perception abilities, MBEA scores revealed that two healthy comparison participants and eight patients (4 RBD and 4 LBD) were amusic (MBEA scores below the cutoff according to their age). An ANOVA with the factor group (LBD patients, RBD patients, comparisons) (see Table 1) revealed a significant main effect (F (2, 57) = 3.38, p = .04), with only the LBD patients having lower MBEA scores than comparisons (p = .022, other p > .064). For PDT, the ANOVA did not reveal a significant main effect of group (F (2, 51) = 2.11, p = .13).

Finally, the patterns of lesions observed for the patients in cortical and subcortical regions were variable across patients, with overall similar localizations of lesions for LBD and RBD patients (Table 2, Fig. 1).

3.2. Musical emotions

Emotion categorization (Fig. 2A). The main effect of group was significant (F (2, 58) = 5.02, p = .0097, partial η^2 = .15). LBD patients had significantly lower scores than comparisons (p = .0028), but no significant difference was found between RBD patients and comparisons (p = .09) or between the two patient groups (p = .23). The main effect of emotion was Significant (F (2.74, 159.02) = 20.195, ϵ = .91, p < .001, partial η^2 = .26), with Joy and Fear being each better recognized than Sadness or Serenity (all p < .001). The interaction of group with emotion was not significant (F (5.48, 159.02) = .76, ϵ = .91, p = .59, partial η^2 = .026).

The correlation between correct emotion categorizations and MBEA scores was significant when pooling data across the three groups (r (58) = -.54, p < .001). A significant correlation was found for the group of LBD patients (r (14) = .57, p = .022) and for comparisons (r (26) = .51, p = .006), but not for RBD patients (r (14) = .35, p = .19) (Fig. 3A). The correlation between correct emotion categorizations and PDT was significant over the three groups (r (52) = -.3, p = .027). No significant correlation was found for RBD patients (r (14) = -.47, p = .1), and for LBD patients (r (14) = .025, p = .94), but the correlation was significant for comparisons (r (26) = -.5, p = .007) (Fig. 3B).

Additional analysis with MBEA covariate.[2] The main effect of group was still nearly significant (F (2, 56) = 3, p = .058, partial η^2 = .097). LBD patients had significantly lower scores than comparisons (p = .05), but no significant difference was found between RBD patients and comparisons (p = .36) or between the two patient groups (p = .32). The main effect of emotion was no longer significant (F (2.77, 155.01) = 1.19, ϵ = .92, p = .32, partial η^2 = .021). The interaction of group with emotion was not significant (F (5.54, 155.01) = .79, ϵ = .92, p = .58, partial η^2 = .026). The main effect of MBEA was significant (F (1, 56) = 15.3, p < .001, partial η^2 = .215). The interaction of MBEA with emotion was not significant (F (2.77, 155.01) = 1.17, ϵ = .92, p = .32, partial η^2 = .02).

Intensity ratings for correct responses (Fig. 2B). The entire range (from 1 to 5) of intensity ratings was covered by the participants, showing that over the groups, the subjective scale was fully used when rating the stimuli. One RBD patient was excluded from the analysis of intensity ratings because for sad musical excerpts, recognition performance was 0%.

The main effect of group was not significant (F (2, 57) = .52, p = .60, partial η^2 = .018). The main effect of emotion was significant (F (2.63, 149.64) = 4.99, ϵ = .88, p = .0024, partial η^2 = .08), with Joy rated as more intense than Sadness, Fear, and Serenity

[2] We performed an additional analysis with PDT as a covariate on 54 participants (6 PDT scores were missing) on categorization scores of musical emotions. This analysis gave similar pattern of results as in the main analysis, and no effect or interaction involving PDT was significant.

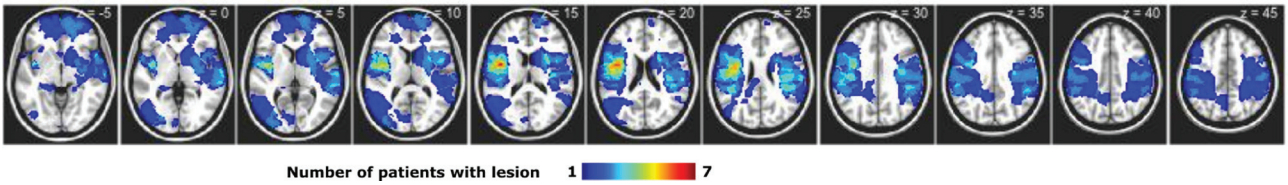

Fig. 1 – Localization of patients' lesions. Overlay of lesions in the patient groups revealed a quite distributed localizations of lesions with similar patterns in left and right hemisphere. Missing data: 1 RBD and 2 LBD patients.

($p < .001$, $p = .011$, and $p = .018$ respectively). The interaction of group with emotion was significant (F (5.25, 149.64) = 2.46, $\epsilon = .88$, $p = .026$, partial $\eta^2 = .079$). RBD patients rated Serenity as more intense than Sadness and Fear (all $p < .003$), whereas no such pattern was observed in the two other groups ($p > .09$). Comparisons rated Joy higher than Sadness and Serenity ($p = .009$ and $p = .004$, respectively). RBD patients had lower intensity ratings for Fear compared to comparisons ($p = .037$), and marginally lower intensity ratings for Fear compared to LBD patients ($p = .10$) (all other $p > .13$).[3]

The correlation between intensity ratings and MBEA scores was not significant over the three groups (r (58) = -.10, $p = .44$) nor in any of the three groups: for RBD patients (r (14) = .37, $p = .16$), for LBD patients (r (14) = -.26, $p = .33$), and for comparisons (r (26) = .10, $p = .61$) (Fig. 3C). The correlation between the intensity ratings and the PDT was not significant over the three groups (r (52) = .14, $p = .33$) nor in any of the three groups: for RBD patients (r (14) = .13, $p = .67$), for LBD patients (r (14) = .45, $p = .12$), and for comparisons (r (26) = -.074, $p = .71$) (Fig. 3D).

Additional analysis with MBEA covariate.[4] The main effect of group was not significant (F (2, 55) = .413, $p = .66$, partial $\eta^2 = .015$). The main effect of emotion was no longer significant (F (2.67, 146.56) = 1.38, $\epsilon = .89$, $p = .25$, partial $\eta^2 = .024$). The interaction of group with emotion was significant (F (4.29, 146.56) = 2.48, $\epsilon = .89$, $p = .046$, partial $\eta^2 = .076$). RBD patients rated Serenity as more intense than Sadness and Fear (all $p < .003$), whereas no such pattern was observed in the two other groups ($p > .3$). Comparisons rated Joy higher than Sadness and Serenity ($p = .009$ and $p = .004$, respectively). RBD patients had lower intensity ratings for Fear compared to comparisons ($p = .037$), and marginally lower intensity ratings for Fear compared to LBD patients ($p = .10$) (all other $p > .13$). The main effect of MBEA was not significant (F (1, 55) = .16, $p = .70$, partial $\eta^2 = .003$). The interaction of MBEA with emotion was not significant (F (2.67, 146.56) = 1.48, $\epsilon = .89$, $p = .23$, partial $\eta^2 = .026$).

[3] An additional ANOVA was performed on all intensity ratings (not only for intensity ratings of the correctly categorized trials). This showed similar results, notably with the main effect of Emotion being significant ($p < .001$) and the interaction between Group and Emotion falling just short of significance ($p = .059$).

[4] We performed an additional analysis with PDT as a covariate on 54 participants (6 PDT scores were missing) on intensity ratings of musical emotions. This analysis gave similar pattern of results as in the main analysis, and no effect or interaction involving PDT was significant.

3.3. *Facial emotions*

Emotion categorization (Fig. 2C). The main effect of group was not significant (F (2, 58) = 1.78, $p = .18$, partial $\eta^2 = .059$). The main effect of emotion was significant (F (2.12, 120.84) = 43.09, $\epsilon = .71$, $p < .001$, partial $\eta^2 = .43$), with Joy and Fear being better recognized than Sadness and Neutrality (all $p < .001$). The interaction between group and emotion was not significant (F (4.24, 120.84) = 1.84, $\epsilon = .71$, $p = .094$, partial $\eta^2 = .061$).

Intensity ratings for correct responses (Fig. 2D). The entire range (from 1 to 5) of intensity ratings was covered by the participants, showing that over the groups, the subjective scale was fully used when rating the stimuli.

The main effect of group was not significant (F (2, 58) = .05, $p = .95$, partial $\eta^2 = .002$). The main effect of emotion was significant (F (1.97, 114.46) = 48.72, $\epsilon = .99$, $p < .001$, partial $\eta^2 = .47$), with Joy rated higher than Sadness and Fear ($p < .001$, and $p = .035$ respectively), and Fear rated higher than Sadness ($p < .001$). The interaction of group with emotion was not significant (F (3.95, 114.46) = 1.012, $\epsilon = .99$, $p = .4$, partial $\eta^2 = .034$).[5]

Data for music and faces material were also analyzed together with a 3 × 4 × 2 ANOVA for emotion categorization with Group (LBD vs RBD patients vs comparisons) as the between-participant factor and Emotion (Joy, Sadness, Fear, Neutrality/Serenity) and Task (Music vs Face) as the within-participant factors. For Intensity ratings, a 3 × 3 × 2 ANOVA was done as Neutrality with Faces material did not have intensity ratings. For correct categorization, the main effect of Task, Emotion and Group were significant ($p < .001$, $p < .001$ and $p = .009$ respectively) as the interactions between Task and Group ($p = .021$), and between Task and Emotion ($p = .017$). Post-hoc revealed that the three participant groups had higher scores for faces than music (all $p < .006$), for music material, comparisons had higher scores that LBD patients ($p = .001$). Post-hoc revealed significant higher scores for faces material compared to music material for Joy, neutrality/Serenity and Fear (all $p < .001$). For intensity ratings, the main effect of Emotion was significant ($p < .001$), as well as the interaction of Task and Emotion ($p < .001$). The triple interaction of Task, Group and Emotion was nearly significant ($p = .057$). For faces material, post-hoc revealed that Sadness was rated lower than Fear and Joy for the three groups (all $p < .04$).

[5] An additional ANOVA was performed on all the intensity ratings (not only for intensity ratings of the correctly categorized trials). This showed similar results, notably with the main effect of Emotion being significant ($p < .001$).

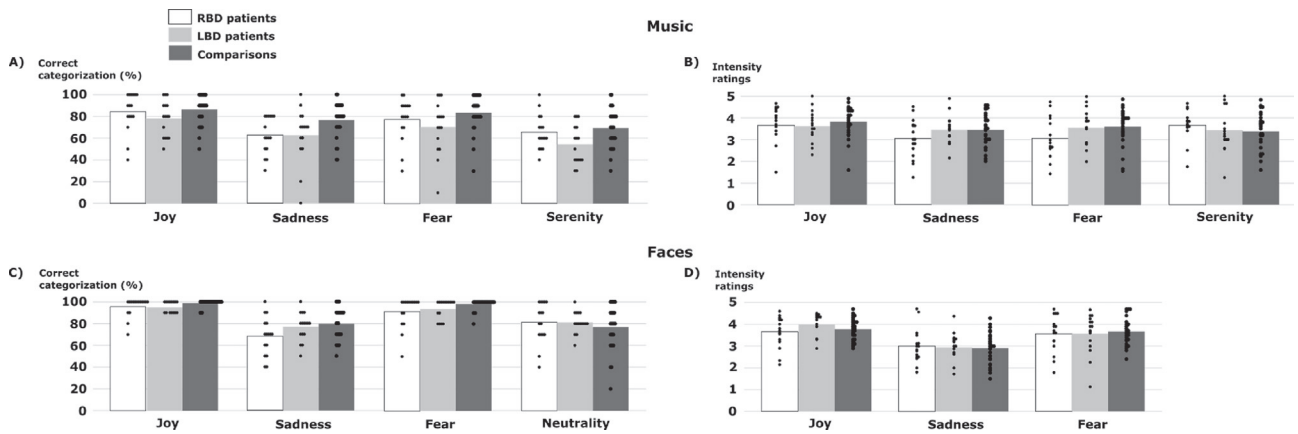

Fig. 2 – Percentage of correct emotion categorization and intensity ratings for music (A and B) and face (C and D) materials in the three groups of participants (RBD patients, LBD patients, comparisons). Bars represent the group means and dots correspond to individual data points. LBD patients had significantly lower correct categorization scores than comparison participants for music material (Panel A). RBD patients had lower intensity ratings for negative emotions in music, a pattern that was not observed in the other two groups (Panel B). All groups showed similar correct categorizations and intensity ratings for faces (Panels C & D).

For music material, post-hoc revealed that Joy was rated higher than Sadness for comparisons and RBD patients (all $p < .038$), interestingly RBD patients also rated Fear lower than Joy ($p = .028$), no such pattern was observed in the other two groups.

3.4. *Testing for potential cross-cultural differences and patient recruitment differences in France and the USA*

3.4.1. *Cross-cultural differences in comparisons participants*

Only effects and interactions involving the factor recruitment site are reported below, effects of emotion mirror the results of the main analyses.

Musical emotion categorization (Fig. 4A). The main effect of site (F (1, 26) = .71, $p = .40$, partial $\eta^2 = .026$) was not significant, neither its interaction with emotion (F (2.42, 62.82) = 1.76, $\epsilon = .81$, $p = .17$, partial $\eta^2 = .063$).

Musical emotion intensity ratings (Fig. 4B). The main effect of site (F (1, 26) = 1.15, $p = .29$, partial $\eta^2 = .042$) was not significant, neither its interaction with emotion (F (2.36, 61.25) = 1.55, $\epsilon = .79$, $p = .22$, partial $\eta^2 = .056$).

Face emotion categorization (Fig. 4C). The main effect of site (F (1,26) = 3.55, $p = .071$, partial $\eta^2 = .12$) did not reach significance, but suggests a slight tendency of Americans comparisons to have better recognition scores compared to French comparisons. The interaction between site and emotion was not significant (F (1.86, 48.42) = .83, $\epsilon = .62$, $p = .43$, partial $\eta^2 = .031$).

Face emotion intensity ratings (Fig. 4D). The main effect of site (F (1, 26) = .11, $p = .74$, partial $\eta^2 = .0043$) was not significant, neither its interaction with emotion (F (1.83, 47.58) = .34, $\epsilon = .92$, $p = .70$, partial $\eta^2 = .013$).

3.4.2. *Patient recruitment across the two sites*

Only effects and interactions involving the site factor are reported below, effects and interactions of emotion and group mirror the results of the main analyses.

Musical emotion categorization (Fig. 5A). The main effect of site was not significant (F (1, 28) = .34, $p = .57$, partial $\eta^2 = .012$), neither its interaction with emotion (F (2.75, 77.03) = .49, $\epsilon = .92$, $p = .67$, partial $\eta^2 = .017$), nor its interaction with lesion-side (F (1.28) = .0029, $p = .96$, partial $\eta^2 < .001$). The three-way interaction of lesion-side, emotion and site was not significant (F (2.75, 77.03) = 2.1, $\epsilon = .92$, $p = .11$, partial $\eta^2 = .07$).

Musical emotion intensity ratings (Fig. 5B). The main effect of site was not significant (F (1, 27) = 1.27, $p = .27$, partial $\eta^2 = .045$), neither its interaction with emotion (F (2.77, 74.71) = 2.46, $\epsilon = .92$, $p = .074$, partial $\eta^2 = .083$), nor its interaction with lesion-side (F (1, 27) = .048, $p = .49$, partial $\eta^2 = .018$). The three-way interaction of lesion-side, emotion and site was not significant (F (2.77, 74.71) = .78, $\epsilon = .92$, $p = .50$, partial $\eta^2 = .028$). The marginal interaction between the effect of site and emotion revealed a slight tendency of French patients to rate higher the intensity of Fear stimuli compared to American patients.

Face emotion categorization (Fig. 5C). The main effect of site was not significant (F (1, 28) = 3.61, $p = .068$, partial $\eta^2 = .11$), neither its interaction with emotion (F (2.25, 63.08) = 1.41, $\epsilon = .75$, $p = .25$, partial $\eta^2 = .048$), nor its interaction with lesion-side (F (1.28) = .044, $p = .84$, partial $\eta^2 = .002$). The three-way interaction of lesion-side, emotion and site was not significant (F (2.25, 63.08) = .18, $\epsilon = .75$, $p = .91$, partial $\eta^2 = .006$).

Face emotion intensity ratings (Fig. 5D). The main effect of site was not significant (F (1, 28) = .61, $p = .44$, partial $\eta^2 = .019$), neither its interaction with emotion (F (1.96, 54.85) = .87, $\epsilon = .98$, $p = .42$, partial $\eta^2 = .030$), nor its interaction with lesion-side (F (1.28) = 2.86, $p = .1$, partial $\eta^2 = .093$). The three-way interaction of lesion-side, emotion and site was not significant (F (1.96, 54.85) = 1.47, $\epsilon = .98$, $p = .24$, partial $\eta^2 = .050$).

According to these results, potential differences of patient recruitment across site cannot be considered as a major source of variability or groups differences observed in our study. It seems that even though the recruitment of patients

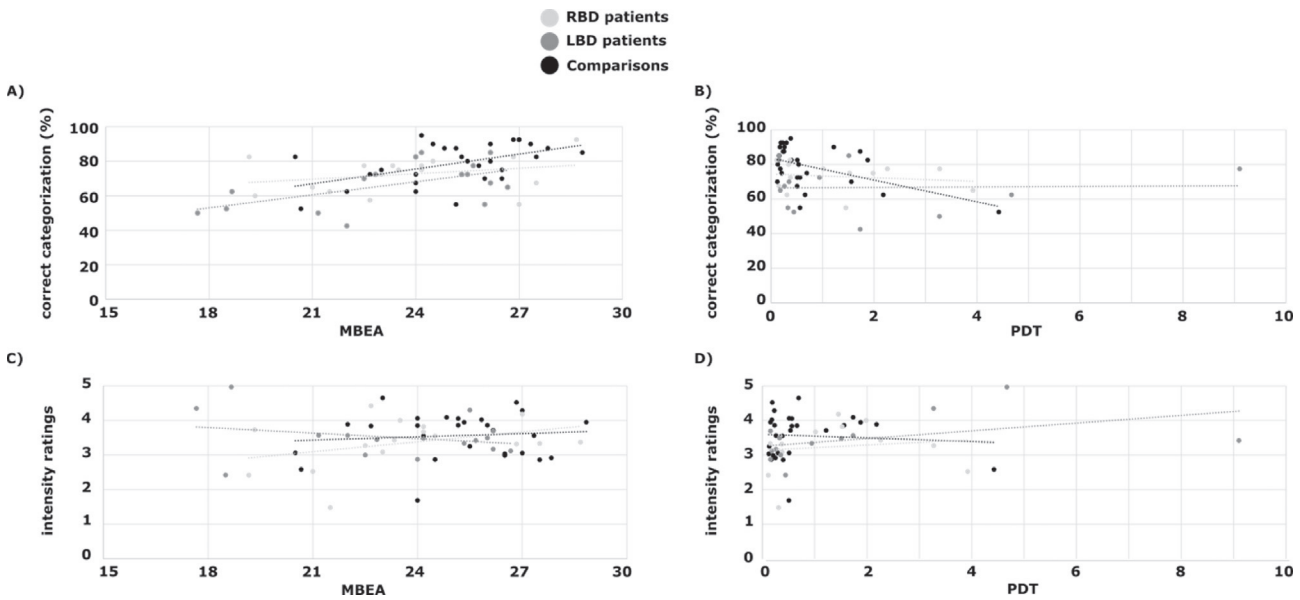

Fig. 3 – Correlations between MBEA and PDT scores and correct categorizations and intensity ratings of music material in the three groups of participants (RBD patients, LBD patients and comparisons). Significant correlation between MBEA score and correct categorization (A) was found for LBD patients (r (14) = .57, p = .022) and for comparisons (r (26) = .51, p = .006). Significant correlation between PDT and correct categorization (B) was found for comparisons (r (26) = -.5, p = .007). No significant correlation was found between intensity ratings and MBEA score (C) and PDT (D), respectively.

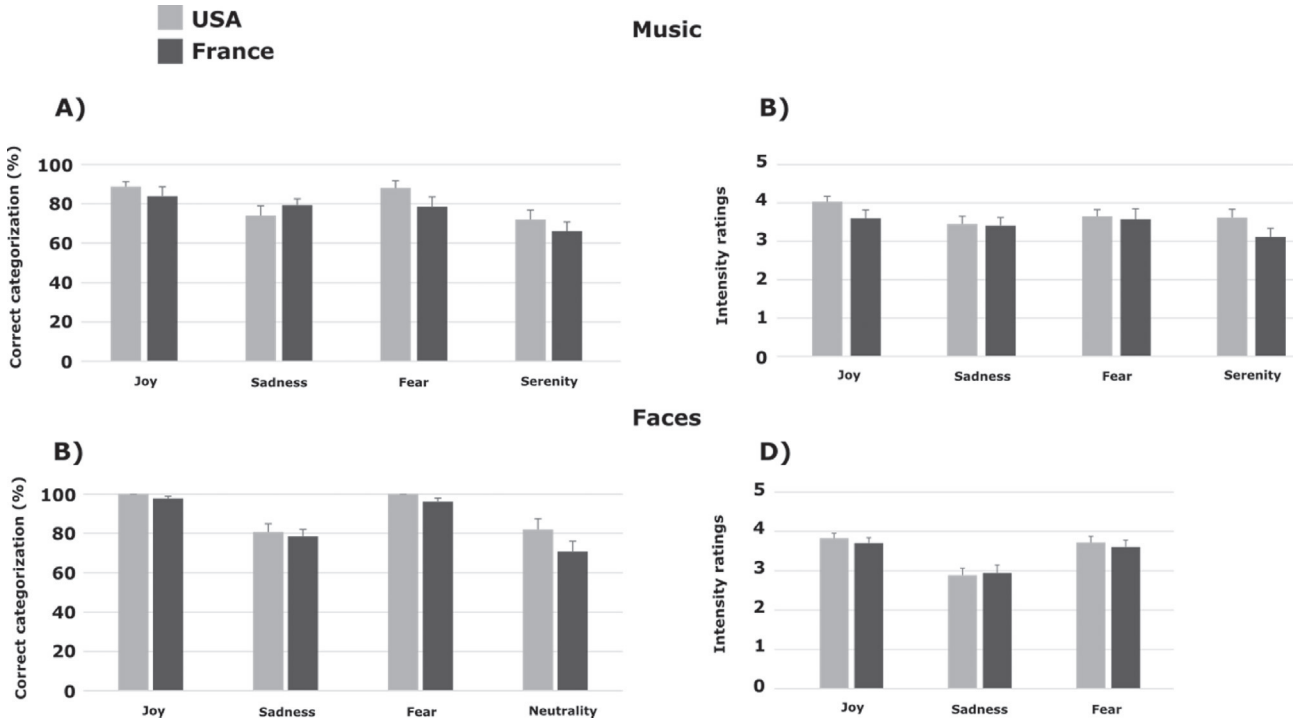

Fig. 4 – Percentages of correct categorization and intensity ratings of comparisons from France and USA, with music (A and B) and face (C and D) materials. No difference was observed between comparisons s from France and USA with both materials, confirming that cross-cultural differences did not influence significantly the results. Error bars indicate the standard error of the mean.

was conducted in two countries, with slightly different inclusion criteria, similar patterns of results were observed on both sites. In conclusion, the results observed with facial and musical emotions are observed across the two western cultures and reflect potential deficits in patients compared to comparisons.

4. Discussion

The present study investigated musical and facial emotion processing after unilateral brain damage. Participants had to categorize the emotion of musical excerpts or faces and rate the intensity of the emotion. Performance in the musical emotion recognition test was significantly lower in LBD patients than comparison participants. RBD patients were not impaired for musical emotion recognition, but rated the emotional intensity of music lower for sadness and fear than for joy and serenity; this difference in intensity ratings was not observed for LBD patients and comparisons. There was no difference for facial emotions (categorization or intensity) between patients and comparison participants, suggesting that the patient groups did not present a general emotion deficit or alteration.

4.1. *Deficits of musical emotion recognition after unilateral brain damage*

Recognition scores of musical emotions revealed a significant deficit in LBD patients compared to comparisons. No significant deficit was observed in RBD patients; note, however, that their performance was numerically in-between that of comparisons and LBD patients. Previous case reports in brain-damaged patients already reported deficits in music emotion recognition associated to various lesions sites (Gosselin et al., 2011, 2007; Griffiths et al., 2004), but no clear association between the lesion site and the deficits has been made. Hence, previous group studies have investigated musical emotion recognition in brain-damaged patients, but focusing up to now only on lesion locations in the mesio-temporal area. The findings of Khalfa et al. (2007, 2008) were in line with our results, notably with stronger impairment of LBD patients than in RBD patients. Note however the slightly different patterns across studies, with one study showing deficit of the LBD patients for sadness and happiness (Khalfa et al., 2008), but another for sadness and anger (Khalfa et al., 2007), and here a more distributed deficit across all emotions in LBD patients. In contrast, Jafari et al. (2017) observed stronger impairment for RBD patients than LBD patients with music material, in particular for sadness and neutrality. Another group study on patients with temporal lobe resection did not find any deficit in LBD and RBD patients in comparison to comparisons for valence and arousal categorizations (Dellacherie et al., 2011). Altogether, the restriction of lesion location in the mesio-temporal area in these studies restricted conclusions. Our study extends the link between potential musical emotion perception deficits and involved brain structures by investigating more various lesion locations than previous studies. It also allows for a comparison between left and right brain damage, with a stronger deficit for emotion recognition associated to left hemisphere damage. Beyond the laterality differences observed here, there were no clear associations between the pattern of musical emotion recognition performance and individual lesion localizations (Table 2). For example, patients showing a deficit at the individual level in musical emotion categorization had a lesion in either parietal, frontal, or temporal cortex. One could argue that there were slightly more LBD patients with a lesion to the insula compared to RBD patients (10 vs 7 patients) that could have influenced the musical emotion recognition results. However, there were no clear association between insula lesion and individual deficit of musical emotion recognition as only half of the LBD patients showing individual deficit also had insula lesion, and only one RBD patient with insula lesion had individual deficit.

Differences in the duration of the used musical material might explain some of the differences observed between previous and our results. Previous studies investigating musical emotions in brain-damaged patients used shorter excerpts of music than we did. In most studies, the stimuli lasted less than 10 s on average (5 s in Dellacherie et al., 2011; 1.5 s in Jafari et al., 2017; 7 s in Khalfa et al., 2008), which limits the number of acoustic cues available to make a decision about the presented emotion. Even though these stimuli might be long enough for comparison participants to detect and identify an emotion (Bigand et al., 2005), they might be too short for patients to make the same judgement. As previously shown in individuals with congenital amusia, the duration of stimuli is essential to allow for extracting a sufficient number of acoustic cues to determine the emotion (Pralus et al., 2019). In the present study, we used musical excerpts of an average duration of 20 s aiming to put participants in the best situation to recognize the emotion. This could explain why we found no deficit in RBD patients. However, the deficit of musical emotion recognition was still present in LBD patients. These results are similar to Khalfa et al. (2007), who also used stimuli that lasted 20 s on average. Moreover, some studies used excerpts played with just one instrument (piano or violin), which could also explain the difference observed between their results and ours (Jafari et al., 2017; Khalfa et al., 2008). Here, we used orchestrated musical extracts to communicate stronger emotions with the use of ecologically valid music, and avoid the potential confound of deficits in the processing of specific timbres (see also Khalfa et al., 2007).

Deficits in musical emotion recognition in LBD patients were not linked to facial emotion recognition deficits in the present study. In contrast to previous studies on facial emotion recognition after brain damage (Borod et al., 2002; Charbonneau et al., 2003; Cheung et al., 2006; Harciarek et al., 2006), no deficit was observed here for patients on the facial task at the group level. Note however, that the facial task was

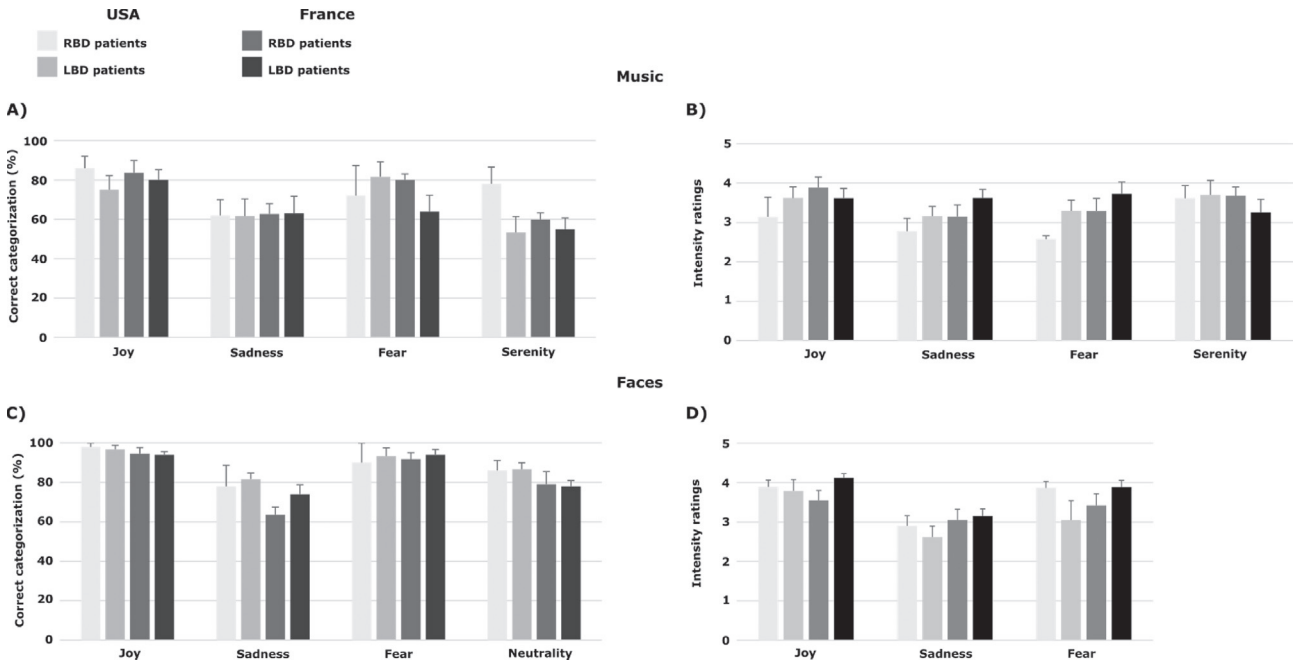

Fig. 5 – Percentages of correct categorization and intensity ratings of patients (RBD and LBD) from France and USA, with music (A and B) and face (C and D) materials. No difference was observed between patients from France and USA with both materials, confirming that selection of patients from both countries did not significantly influence the results. Error bars indicate the standard error of the mean.

easier than the musical task, as revealed by the higher scores obtained by comparison participants. This suggests that the Right Hemisphere Hypothesis previously supported by facial material (Borod, Bloom, Brickman, Nakhutina, & Curko, 2002; Charbonneau et al., 2003; Harciarek et al., 2006; Kucharska-Pietura, Phillips, Gernand, & David, 2003; Tippett et al., 2018) may be specific to facial material and not generalized to all emotions recognition. These results with facial material also confirmed that despite the depression scores below the cut-off in some patients, they did not have a general emotional deficit that could have influenced the results with music material.

Three RBD patients had a brain damage in the basal ganglia, which could have influenced the group results. Another study on facial emotion recognition showed that patients with localized basal ganglia damage performed significantly worse in recognizing negative emotions than comparisons (anger, disgust and fear) (Cheung, Lee, Yip, King, & Li, 2006). However, at the group level, we did not observe any difference between the three groups for facial emotion recognition.

4.2. *Links between musical emotion recognition and music perception*

Over the three participant groups, percentage of correct categorization of musical emotions correlated positively with the MBEA mean score. This was also the case for LBD patients and comparisons, but not for RBD patients. LBD patients had a lower MBEA mean score compared to comparisons. Moreover, when MBEA was considered in the categorization scores analysis, we demonstrated that the effect of MBEA was indeed significant, demonstrating a potential effect of the deficit of musical perception in LBD patients on musical emotion recognition results. These results of LBD patients are in agreement with a study on congenital amusia (Lévêque et al., 2018) showing that congenital amusic individuals (diagnosed by low MBEA scores) were impaired in musical emotion categorization in comparison to comparison participants. These findings reveal that some of the participants could have a global deficit in evaluating musical stimuli (Särkämö et al., 2009; Tillmann, Albouy, & Caclin, 2015). This might also reflect deficits in more general cognitive abilities required by the MBEA (Särkämö et al., 2009, 2010). Indeed, three LBD patients had also a MMSE score below the cut-off which could have influenced the MBEA results. However, this medium cognitive deficit could not be the only cause of musical emotion perception deficit as LBD patients were not impaired for facial emotion perception.

Based on the present group-level results, we can argue that cognitive and perceptual musical abilities are important for explicitly recognizing musical emotions, as the LBD patients showed decreased MBEA scores as well as deficits in musical

emotion recognition. However other parameters must be involved in recognizing musical emotions, as patients can have acquired amusia without deficit in categorization of musical emotions (see patient F11, also in Hirel et al., 2014) or participants with congenital amusia can demonstrate preserved sensitivity to emotional music (Gosselin, Paquette, & Peretz, 2015). The variety of profiles observed among the present patient sample are in keeping with the hypothesis of (at least partly) separate processes for music perception and emotion (Peretz et al., 1998; Satoh et al., 2011; Stewart, von Kriegstein, Warren, & Griffiths, 2006). Furthermore, we did not observe any link between emotion intensity ratings and perceptual musical abilities.

4.3. *A deficit of valence processing in musical emotions in RBD patients*

For the musical materials, RBD patients did not show any deficit on musical emotion recognition, but rated the emotional intensity of music lower for sadness and fear than for joy and serenity, in agreement with the valence hypothesis. This pattern of ratings was not observed in the two other groups (LBD patients and comparisons). Previous group studies on mesio-temporal lobe damaged patients also provided data in line with the validity of the valence hypothesis using a task of musical emotion recognition (Jafari et al., 2017; Khalfa et al., 2007, 2008). The present results further support this hypothesis based on patient groups with more diverse lesion locations and on intensity ratings of musical emotions.

4.4. *Clinical interest of assessing musical emotions*

In the present results, it is noteworthy that RBD patients did not show any deficit in emotion categorization, whereas they exhibited an abnormal pattern of intensity ratings of musical emotions. This pattern suggests that conceptual knowledge about emotion categories can persist even when the intensity of emotions is abnormally perceived. Intensity ratings may reflect more implicit representation of the emotion and could be linked to what emotions the listener really feels (Hirel et al., 2014; Lévêque et al., 2018). For congenital amusic participants, this paradigm has revealed a reverse pattern compared to the present study, with preserved implicit capacities to process musical emotions (i.e., with preserved intensity ratings), but impairments in the classical explicit categorization test (Lévêque et al., 2018; see also; Tillmann, Lalitte, Albouy, Caclin, & Bigand, 2016). Intensity ratings can be considered as an implicit investigation method as no verbal categorization of a given emotion and only a weak internal representation of the stimulus is necessary to provide a judgement. In the present study, intensity ratings allowed revealing deficits in patients that could not be detected with the recognition paradigm. Thus, intensity ratings, in combination with explicit recognition measures, could allow building a sensitive test to detect possible emotion perception abnormalities in clinical settings, even if a patient is unaware of this deficit (Stewart et al., 2006; Tillmann et al., 2016). Moreover, this paradigm reveals the distinction between cognitive intentional process of emotion recognition, and the emotional experience of music in unilateral brain-damaged patients, as it was already suggested in healthy participants and unilateral brain-damaged patients with facial and vocal stimuli (Abbott et al., 2013; Borod et al., 2002; Demaree et al., 2005).

5. Conclusion

The present study revealed two major patterns of potential deficits in musical emotion processing after brain damage. Our findings reveal a specific deficit for musical emotion categorization in LBD patients, whereas intensity ratings showed that right brain-damaged patients underrated negative valence emotions (compared to left brain-damaged patients and comparisons). Intensity rating data were thus compatible with the valence hypothesis, and the overall data pattern refines the distinction between the roles of the two hemispheres: the right hemisphere seems to be important to experience emotions, in particular negative emotions, whereas the left hemisphere seems to be more strongly involved in recognizing emotions at an explicit level. This hemispheric differentiation extends beyond the mesio-temporal structures of the brain, which were the focus in previous musical emotion studies with brain-damaged patients.

Author contribution

Agathe Pralus: Formal analysis, Investigation, Writing-Original draft
Amy Belfi: Investigation, Conceptualization
Catherine Hirel Investigation, Conceptualization
Yohana Lévêque: Resources
Lesly Fornoni: Investigation
Emmanuel Bigand: Resources
Julien Jung: Resources
Daniel Tranel: Supervision, Funding acquisition
Norbert Nighoghossian: Resources
Barbara Tillmann: Supervision, Funding acquisition, Project administration
Anne Caclin: Supervision, Funding acquisition, Project administration

Declaration of Competing Interest

The authors report no competing interests.

Acknowledgments

We thank the patients and comparison participants for their time and participation to the study. We thank Roxane Hoyer for help with the analyses of neuropsychological data.

This work was supported by grants from Agence Nationale de la Recherche of the French Ministry of Research ANR-11-BSH2-001-01 to BT and AC, and ANR-14-CE30-0001-01 to AC.

CH was funded by an "Année Recherche" grant from "ARS (Agence Régionale de Santé) Rhône-Alpes". This work was conducted in the framework of the LABEX CELYA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060) and of the LabEx Cortex ("Construction, Function and 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.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2020.05.015.

REFERENCES

Abbott, J. D., Cumming, G., Fidler, F., & Lindell, A. K. (2013). The perception of positive and negative facial expressions in unilateral brain-damaged patients: A meta-analysis. *Laterality, 18*(4), 437–459. https://doi.org/10.1080/1357650X.2012.703206.

Adolphs, R., Jansari, A., & Tranel, D. (2001). Hemispheric perception of emotional valence from facial expressions. *Neuropsychology, 15*(4), 516–524.

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*(3), 396–404.

Altenmüller, E., Schürmann, K., Lim, V. K., & Parlitz, D. (2002). Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia, 40*(13), 2242–2256. https://doi.org/10.1016/s0028-3932(02)00107-0.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience, 10*(3), 399–407. https://doi.org/10.1093/scan/nsu067.

Belfi, A. M., Evans, E., Heskje, J., Bruss, J., & Tranel, D. (2017). Musical anhedonia after focal brain damage. *Neuropsychologia, 97*, 29–37. https://doi.org/10.1016/j.neuropsychologia.2017.01.030.

Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences, 1060*, 429–437. https://doi.org/10.1196/annals.1360.036.

Borod, J. C. (1992). Interhemispheric and intrahemispheric control of emotion: A focus on unilateral brain damage. *Journal of Consulting and Clinical Psychology, 60*(3), 339–348. https://doi.org/10.1037/0022-006X.60.3.339.

Borod, J. C., Bloom, R. L., Brickman, A. M., Nakhutina, L., & Curko, E. A. (2002). Emotional processing deficits in individuals with unilateral brain damage. *Applied Neuropsychology, 9*(1), 23–36. https://doi.org/10.1207/S15324826AN0901_4.

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*. https://doi.org/10.1007/s10072-019-04175-w.

Braun, M., Traue, H. C., Frisch, S., Deighton, R. M., & Kessler, H. (2005). Emotion recognition in stroke patients with left and right hemispheric lesion: Results with a new instrument-the FEEL test. *Brain and Cognition, 58*(2), 193–201. https://doi.org/10.1016/j.bandc.2004.11.003.

Charbonneau, S., Scherzer, B. P., Aspirot, D., & Cohen, H. (2003). Perception and production of facial and prosodic emotions by chronic CVA patients. *Neuropsychologia, 41*(5), 605–613. https://doi.org/10.1016/S0028-3932(02)00202-6.

Cheung, C. C. Y., Lee, T. M. C., Yip, J. T. H., King, K. E., & Li, L. S. W. (2006). The differential effects of thalamus and basal ganglia on facial emotion recognition. *Brain and Cognition, 61*(3), 262–268. https://doi.org/10.1016/j.bandc.2006.01.008.

Dellacherie, D., Bigand, E., Molin, P., Baulac, M., & Samson, S. (2011). Multidimensional scaling of emotional responses to music in patients with temporal lobe resection. *Cortex. a Journal Devoted to the Study of the Nervous System and Behavior, 47*(9), 1107–1115. https://doi.org/10.1016/j.cortex.2011.05.007.

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: Historical roots and a future incorporating "dominance". *Behavioral and Cognitive Neuroscience Reviews, 4*(1), 3–20. https://doi.org/10.1177/1534582305276837.

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2014). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese pygmies. *Frontiers in Psychology, 5*, 1341. https://doi.org/10.3389/fpsyg.2014.01341.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Consulting psychologists Press.

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review, 17*(3), 335–341. https://doi.org/10.3758/PBR.17.3.335.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198.

Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in Neurobiology, 123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003.

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 71*, 171–182. https://doi.org/10.1016/j.cortex.2015.06.022.

Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 47*(9), 1116–1125. https://doi.org/10.1016/j.cortex.2011.05.012.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia, 45*(2), 236–244. https://doi.org/10.1016/j.neuropsychologia.2006.07.012.

Gosselin, N., Samson, S., Adolphs, R., Noulhiane, M., Roy, M., Hasboun, D., et al. (2006). Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. *Brain: a Journal of Neurology, 129*(Pt 10), 2585–2592. https://doi.org/10.1093/brain/awl240.

Griffiths, T. D., Warren, J. D., Dean, J. L., & Howard, D. (2004). "When the feeling's gone": A selective loss of musical emotion. *Journal of Neurology, Neurosurgery, and Psychiatry, 75*(2), 344–345. https://doi.org/10.1136/jnnp.2003.015586.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*, 56–62. https://doi.org/10.1136/jnnp.23.1.56.

Harciarek, M., & Heilman, K. M. (2009). The contribution of anterior and posterior regions of the right hemisphere to the recognition of emotional faces. *Journal of Clinical and*

Experimental Neuropsychology, 31(3), 322–330. https://doi.org/10.1080/13803390802119930.
Harciarek, M., Heilman, K. M., & Jodzio, K. (2006). Defective comprehension of emotional faces and prosody as a result of right hemisphere stroke: Modality versus emotion-type specificity. *Journal of the International Neuropsychological Society: JINS, 12*(6), 774–781. https://doi.org/10.1017/S1355617706061121.
Hartman, D. E. (2009). Wechsler Adult intelligence scale IV (WAIS IV): Return of the gold standard. *Applied Neuropsychology, 16*(1), 85–87. https://doi.org/10.1080/09084280802644466.
Hirel, C., Lévêque, Y., Deiana, G., Richard, N., Cho, T.-H., Mechtouff, L., et al. (2014). Acquired amusia and musical anhedonia. *Revue Neurologique, 170*(8–9), 536–540. https://doi.org/10.1016/j.neurol.2014.03.015.
Jafari, Z., Esmaili, M., Delbari, A., Mehrpour, M., & Mohajerani, M. H. (2017). Post-stroke acquired amusia: A comparison between right- and left-brain hemispheric damages. *Neurorehabilitation, 40*(2), 233–241. https://doi.org/10.3233/NRE-161408.
Joanette, Ska, & Côté. (2004). *Protocole MEC-P - protocole Montréal d'évaluation de la communication de poche*. Cultura.Com. http://www.cultura.com/protocole-mec-p-9782362350030.html.
Johnsen, E. L., Tranel, D., Lutgendorf, S., & Adolphs, R. (2009). A neuroanatomical dissociation for emotion induced by music. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 72*(1), 24–33. https://doi.org/10.1016/j.ijpsycho.2008.03.011.
Kaplan, E., Goodglass, H., Weintraub, S., & Goodglass, H. (1983). Boston naming test. *Lea & Febiger*.
Khalfa, Delbe, C., Bigand, E., Reynaud, E., Chauvel, P., & Liéégeois-Chauvel, C. (2007). Positive and negative music recognition reveals a specialization of mesio-temporal structures in epileptic patients. *Music Perception: An Interdisciplinary Journal, 25*(4), 295–302. https://doi.org/10.1525/mp.2008.25.4.295.
Khalfa, Guye, M., Peretz, I., Chapon, F., Girard, N., Chauvel, P., et al. (2008). Evidence of lateralized anteromedial temporal structures involvement in musical emotion processing. *Neuropsychologia, 46*(10), 2485–2493. https://doi.org/10.1016/j.neuropsychologia.2008.04.009.
Khalfa, S., Schon, D., Anton, J.-L., & Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport, 16*(18), 1981–1984.
Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., et al. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *Neuroimage, 81*, 49–60. https://doi.org/10.1016/j.neuroimage.2013.05.008.
Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia, 41*(8), 1082–1090.
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*. https://doi.org/10.1037/neu0000461.
Liégeois-Chauvel, C., Bénar, C., Krieg, J., Delbé, C., Chauvel, P., Giusiano, B., et al. (2014). How functional coupling between the auditory cortex and the amygdala induces musical emotion: A single case study. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 60*, 82–93. https://doi.org/10.1016/j.cortex.2014.06.002.
Loonstra, A. S., Tarlow, A. R., & Sellers, A. H. (2001). COWAT metanorms across age, education, and gender. *Applied Neuropsychology, 8*(3), 161–166. https://doi.org/10.1207/S15324826AN0803_5.
Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 12*(4), 313–324. https://doi.org/10.1016/S0010-9452(76)80035-4.
Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930.
Nijboer, T. C. W., & Jellema, T. (2012). Unequal impairment in the recognition of positive and negative emotions after right hemisphere lesions: A left hemisphere bias for happy faces. *Journal of Neuropsychology, 6*(1), 79–93. https://doi.org/10.1111/j.1748-6653.2011.02007.x.
Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex. *Annals of the New York academy of sciences*. https://doi.org/10.1111/nyas.13666.
Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain: a Journal of Neurology, 113*(Pt 4), 1185–1205.
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*, 58–75.
Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68*(2), 111–141. https://doi.org/10.1016/S0010-0277(98)00043-2.
Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., et al. (2019). *Emotional prosody in congenital amusia: Impaired and spared processes* (p. 107234). Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2019.107234.
Sackeim, H. A., Greenberg, M. S., Weiman, A. L., Gur, R. C., Hungerbuhler, J. P., & Geschwind, N. (1982). Hemispheric asymmetry in the expression of positive and negative emotions. Neurologic evidence. *Archives of Neurology, 39*(4), 210–218.
Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., et al. (2009). Cognitive deficits associated with acquired amusia after stroke: A neuropsychological follow-up study. *Neuropsychologia, 47*(12), 2642–2651. https://doi.org/10.1016/j.neuropsychologia.2009.05.015.
Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., et al. (2010). Auditory and cognitive deficits associated with acquired amusia after stroke: A magnetoencephalography and neuropsychological follow-up study. *Plos One, 5*(12). https://doi.org/10.1371/journal.pone.0015157.
Satoh, M., Kato, N., Tabei, K.-I., Nakano, C., Abe, M., Fujita, R., et al. (2016). A case of musical anhedonia due to right putaminal hemorrhage: A disconnection syndrome between the auditory cortex and insula. *Neurocase, 22*(6), 518–525. https://doi.org/10.1080/13554794.2016.1264609.
Satoh, M., Nakase, T., Nagata, K., & Tomimoto, H. (2011). Musical anhedonia: Selective loss of emotional experience in listening to music. *Neurocase, 17*(5), 410–417. https://doi.org/10.1080/13554794.2010.532139.
Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*(3), 239–251. https://doi.org/10.1080/0929821042000317822.
Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain: a Journal of Neurology, 129*(Pt 10), 2533–2553. https://doi.org/10.1093/brain/awl171.
Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handbook of Clinical Neurology, 129*, 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0.
Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks.

Neuropsychologia, 85, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027.

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.

Tippett, D. C., Godin, B. R., Oishi, K., Oishi, K., Davis, C., Gomez, Y., et al. (2018). Impaired recognition of emotional faces after stroke involving right amygdala or insula. *Seminars in Speech and Language, 39*(1), 87–100. https://doi.org/10.1055/s-0037-1608859.

Tsang, C. D., Trainor, L. J., Santesso, D. L., Tasker, S. L., & Schmidt, L. A. (2001). Frontal EEG responses as a function of affective musical features. *Annals of the New York Academy of Sciences, 930*(1), 439–442. https://doi.org/10.1111/j.1749-6632.2001.tb05764.x.

Yuvaraj, R., Murugappan, M., Norlinah, M. I., Sundaraj, K., & Khairiyah, M. (2013). Review of emotion recognition in stroke patients. *Dementia and Geriatric Cognitive Disorders, 36*(3–4), 179–196. https://doi.org/10.1159/000353440.

Supplementary material

FRENCH PATIENTS	NEUPSYCHOLOGICAL TEST	NORM CUT-OFF	RBD PATIENTS (N=11)	LBD PATIENTS (N=10)	*P-VALUE (GROUP COMPARISON)*	RBD PATIENTS OUT OF THE NORM	LBD PATIENTS OUT OF THE NORM
GENERAL COGNITIVE FUNCTIONNING	*MMSE**	24	27.8 +/-1.83 (25-30)	27.3 +/- 3.09 (23-30)	*0.88*	0	3
VERBAL ABILITIES	*Lexical fluencies**	8.09 (<65 years) 5.18 (>65 years)	17.1 +/-7.81 (3-25)	16.3 +/-7.96 (6-28)	*0.97*	2	1
	*Categorical fluencies**	20.46 (<65 years) 10.08 (>65 years)	25 +/-6.47 (20-36)	26.5 +/-7.98 (15-35)	*0.62*	1	0
	Emotional Prosody (MEC)	7 (< 64 years) 5 (> 64 years)	10 +/-2.16 (6-12)	10.8 +/-1.81 (7-12)	*0.40*	0	0
	Linguistic Prosody (MEC)	4	10.1 +/-2.3 (5-12)	9.9 +/-2.38 (5-12)	*0.91*	0	0
OTHERS	*Depression scale**	7	12.6 +/-10.11 (3-31)	19.3 +/-10.23 (5-35)	*0.20*	6	7

Table S1*: Neuropsychological data for French patients. Standard deviations are indicated and minimum-maximum values are indicated in the parentheses. Group comparison were done Mann-Whitney tests. *Missing data for one RBD patient. MMSE: Mini-Mental State Examination. MEC: Montreal Evaluation of Communication, test for the recognition of the emotion conveyed by a narrator in sentences (emotional prosody) and for the recognition of question or affirmation in a pronounced sentence (linguistic prosody).*

AMERICAN PATIENTS	NEUPSYCHOLOGICAL TEST	NORM CUT-OFF	RBD PATIENTS (N=5)	LBD PATIENTS (N=6)	P-VALUE (GROUP COMPARISON)	RBD PATIENTS OUT OF THE NORM	LBD PATIENTS OUT OF THE NORM
GENERAL COGNITIVE FUNCTIONNING	*WAIS_FSIQ*	100 +/-30	101.5 +/-18.6 (79-127)	111.8 +/-12.4 (99-127)	0.41	0	0
	WAIS_WMI	100 +/-30	98.8 +/-21.1 (71-119)	103 +/-4.4 (97-108)	0.93	0	0
	WAIS_PSI	100 +/-30	106.8 +/-22.4 (81-146)	101.4 +/14.9 (92-127)	0.79	0	0
	WCST_PE	15.15+/-9.68 (< 60) 14.27 +/-8.37 (60-69) 19.54 +/-11.06 (> 70)	13.3 +/-12.4 (3-37)	8.4 +/-6.7 (4-20)	0.52	1	0
	WCST_CAT	4.61 +/-1.9 (< 60) 5.13 +/-1.43 (60-69) 4.14 +/-1.96 (> 70)	5.5 +/-1.2 (3-6)	6 +/-0	0.71	0	0
VERBAL ABILITIES	*AVLT*	8.7 +/-3 (< 60) 6.8 +/-3.7 (60-69) 5.6 +/- 2.6 (> 69)	10.2 +/-3.3 (5-14)	11 +/-2.9 (7-14)	0.71	1	0
	BNT	56.1 +/- 9.27	58.2 +/-1.9 (55-60)	55.4 +/-4.4 (48-59)	0.23	0	0
	COWA	34.24 +/-12.48 (<60) 32.31 +/-12.7 (60-79)	39.7 +/-8.9 (31-56)	44 +/-11.8 (28-57)	1	0	0

Table S2*: Neuropsychological data for American participants. Standard deviations are indicated and minimum-maximum values are indicated in the parentheses. Group comparisons were done with Mann Whitney tests. WAIS: Wechsler Adult Intelligence Scale for full-scale IQ (FSIQ), working memory (WMI) and processing speed (PSI). WCST: Wisconsin Card Sorting Task: perseverative errors (PE) and categories completed (CAT). AVLT: an auditory-verbal learning test. BNT: Boston Naming Test. COWA: Controlled Oral Word Association.*

Excerpt	*Composer*	*Piece*
Serenity1	W. F. Bach	Duet No. 4 for 2 Flutes in F Majeur Fk 57.II Lamentabile
Serenity2	J. Brahms	Concerto for Violin and Orchestra in D Major, Op 77, Second Movement: Adagio
Serenity3	C. Franck	Violin Sonata, 4th Movement: Allegretto Pocco Mosso
Serenity4	J. Haydn	Sinfonia Concdrtante in B-Flat Major, Hob. I105: II Andante
Serenity5	R. Strauss	Don Quixote, Finale (Sehr Ruhig)
Serenity6	D. Scarlatti	Keyboard sonata in A Major, K.208 (interpreted on the guitarre)
Serenity7	A. Scarlatti	Lamentazioni Per la Settimana Santa, Motets
Serenity8	L. Beethoven	Piano Sonata No 16 in G Major, Op 31. No.1 II. Adagio grazioso
Serenity9	J. Haydn	Flute Trio No. 6 in D Major Hob. IV-6, I. Adagio cantabile
Serenity10	W. A. Mozart	Synphony No. 24 in B-Flat Major, K. 182: II. Andantino Grazioso
Anger/Fear1	F. Chopin	Prelude Op. 28, no.22
Anger/Fear2	G. Holst	Les planètes: Mars
Anger/Fear3	F. Liszt	Poème symphonique no.2 (Tasso Lamento & Triomfo)
Anger/Fear4	R. Strauss	Tod and Verklarung
Anger/Fear5	JF Rebel	Les éléments
Anger/Fear6	A. Schoenberg	Erwartung, Op 17, Scene IV
Anger/Fear7	D. Shostakovitch	Trio 2, I. Andante: Moderato
Anger/Fear8	F. Liszt	Totentanz.
Anger/Fear9	P.I. Tchaïkovski	Symphony Pathétique, 1st movement
Anger/Fear10	R. Strauss	Don Quixote, op. 35
Joy1	W. F. Bach	Bach: Duet No. 6 for 2 Flutes in G Major, FK 69, I Allegro ma non troppo
Joy2	L v. Beethoven	Piano Sonata No. 32 in C Minor, Op. 111: II Arietta
Joy3	L v. Beethoven	Symphony No. 7, In A Major, Op 92: I Poco Sostenuto Vivace
Joy4	J. Brahms	Trio pour Piano, Violon et cor in E-Flat Major, op 40: Scherzo

Joy5	F. Liszt	Poême symphonique "Les préludes"
Joy6	F. Mendelssohn	Symphony No. 4 in A Major, Op 90 Italian, I. Allegro Vivace
Joy7	J. S. Bach	French Suite No. 6, BWV 820
Joy8	I. Stravinsky	Petrouchka- Scene 1: First Tableau
Joy9	J. S. Bach	Brandenburg Concerto No. 3 in G Major BWV 1048, II. Allegro
Joy10	F. Liszt	Konzert fuer Klavier und Orchester, Nr. 1 Es-Dur III.Allegro
Sadness1	P.I. Tchaïkovski	Symphony Pathétique, 1st movement
Sadness2	G. Mahler	Symphony No8 in E-flat, part 2, final scene from Goethe's Faust
Sadness3	W. A. Mozart	Dissonance Quartet in C Major Adagio
Sadness4	P.I. Tchaïkovski	Symphony Pathétique, 4th movement
Sadness5	S. Rachmaninov	Piano Concerto no. 3 in D Minor, Op 30 II. Intermezzo
Sadness6	D. Shostakovitch	Symphony 15, Adagio
Sadness7	D. Shostakovitch	Trio No. 2 in E Minor , III Largo
Sadness8	P.I. Tchaïkovski	5th symphony, 1st mouvement
Sadness9	P.I. Tchaïkovski	Symphony, No. 5 in E Minor, Op 64, Andate Cantabile
Sadness10	R. Wagner	Tristan, Act 3, Prelude

Table S3*: Musical excerpts selected from the Western classical repertoire. All excerpts were orchestrated instrumental stimuli, without voice, lasted 20 seconds, and were aimed to be representative of four emotions in real recordings. In this selection of stimuli, ten excerpts related to joy, ten to sadness, ten to fear/, ten to serenity.*

Part 2: Boosting non-verbal auditory processing: a new training strategy

VI. Rehabilitation in patients with cochlear implants.

A. Objectives

Non-verbal auditory perception can be altered in the case of peripheral hearing loss. As we have seen in part one-II-C, hearing loss can be partially restored by one or two cochlear implant(s). However, auditory perception in CI users remains particularly impaired for non-verbal auditory cues. Several studies have reported hearing-in-noise difficulties in CI users but also music perception and emotional prosody recognition deficits. In Study 4, we aimed to design a new short assessment tool to test specifically some aspects of non-verbal auditory perception in CI users and normal-hearing (NH) participants with full and vocoded sounds. This testing battery will give a rapid overview of non-verbal abilities in an individual and will help for further research thanks to its implementation on touch tablet. To design this battery, we used in part previous tasks already tested on congenital amusics as they also have a pitch-related deficit.

Moreover, we saw that multisensory integration can be enhanced following a sensory loss. This is why, we could suggest that audiovisual integration could be particularly efficient in CI users. In prevision for the design of a new rehabilitation program for CI users (Study 5), we wanted to test whether visual cues could improve pitch processing for CI users as well as NH processing of full and vocoded signals. To do so, we used both informative and non-informative visual cues in three non-verbal auditory tasks of the testing battery.

Based on the results of Study 4, a new rehabilitation program for non-verbal auditory cognition was designed in Study 5. We focused the learning strategy on audiovisual interaction to benefit from multisensory integration. Indeed, Study 4 demonstrated that audiovisual interaction could enhance perception in NH participants but even more so in CI users. To better characterize the brain networks involved in the learning during the multisensory training, we recorded magnetoencephalography in control participants. As pitch perception and pitch memory are known to involve fronto-temporal cortical networks, we targeted these regions during our recordings. Here, we present the design of the

experiment as well as preliminary results on three control participants for behavioral tasks only. Due to the sanitary conditions in France, all experiments had to be stopped and delayed our recordings.

A. Study 4: Rapid assessment of non-verbal auditory perception in normal-hearing participants and cochlear implant users

Rapid assessment of non-verbal auditory perception in normal-hearing participants and cochlear implant users

Pralus A. [1,2], Hermann R. [1,2,3], Cholvy F. [1,2], Aguera P.E. [1,2], Moulin A. [1,2], Barone P. [4], Grimault N. [1,2], Truy E. [1,2,3], Tillmann B.*[1,2], Caclin A.*[1,2]

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

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

[3] ENT Department, Hôpital Edouard Herriot, Hospices Civils de Lyon, Lyon, F-69000, France

[4] Brain & Cognition Research Center (CerCo); CNRS, UMR 5549; University of Toulouse Paul Sabatier; Toulouse, France.

* Equally contributing authors.

Corresponding author: agathe.pralus@inserm.fr CRNL, 95 Bd Pinel 69675 Bron cedex, France, +33 4 72 13 89 22

Conflicts of Interest: none

Source of Funding: This work was supported by the ‘‘Fédération pour la Recherche sur le Cerveau (FRC) ”. This work was conducted in the framework of the LabEx CeLyA (‘‘Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) and of the LabEx Cortex (‘‘Construction, Function and 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).

Highlights

- **A battery of five listening tests was used to assess non-verbal auditory perception**
- **CI users show deficits in pitch discrimination, emotional prosody, and streaming**
- **Similar deficits are observed in NH listeners with vocoded sounds**
- **Visual cues can enhance CI users' performance in pitch perception tasks**

Abstract

Objectives. As a multi-faceted ability, auditory cognition includes speech processing, music processing, and auditory scene analysis. In case of profound hearing loss, cochlear implants (CI) allow for the restoration of hearing. Despite CI's advantages for speech perception, CI users often experience difficulties for hearing speech in noise, understanding prosody, and music perception. **Design.** Aiming to better characterize non-verbal auditory perception in CI users, we developed a new short battery with five listening tests, each assessing a specific component of non-verbal auditory perception, including pitch processing. These five tests measured pitch change detection, pitch direction identification, pitch short-term memory, auditory stream segregation, and emotional prosody recognition with related intensity ratings. As pitch processing can be improved with visual cues, the three pitch tests were implemented with and without visual cues. We tested 10 Normal-Hearing (NH) participants with original and vocoded sounds, and 10 post-lingually deaf CI users. **Results.** With vocoded sounds, NH participants had reduced emotion recognition and pitch perception (in comparison to original signals), in particular in the pitch change detection task. CI users performed less well than NH participants with original sounds, in particular, for emotion recognition and pitch change detection, and their performance level were similar to NH participants with vocoded sounds. In addition, CI users showed decreased streaming capacity compared to NH participants. For pitch change detection, CI users benefited from uninformative visual cues, leading to enhanced performance in comparison to auditory information only. **Conclusions.** Overall, this battery allows for a rapid detection of non-verbal auditory perception deficits, using the same kind of stimuli for all pitch tests with frequency roving. This battery can be used in CI users, and in NH participants with normal and vocoded sounds. The current findings also open new perspectives

for the training of pitch perception with visual cues, in particular for populations with reduced hearing capacities.

Keywords: hearing loss, pitch perception, auditory scene analysis, prosody, audiovisual integration

1. Introduction

Auditory cognition contributes to perceive our environment and to react properly to external stimuli. Communication via language is essential, but non-verbal auditory cognition also plays a primary role for the understanding of perceived stimuli, including prosody. Indeed, in speech communication, prosody perception is essential to understand the intentions and emotions of the speaker (Lolli et al., 2015). Moreover, non-verbal auditory abilities allow detecting auditory signals in noise as well as better perceiving and analyzing the auditory scene (Bregman, 1994). Music perception also relies on our abilities to process and determine melody and harmony in the non-verbal auditory signal (McAdams, 1989).

Non-verbal auditory perception can be altered in the case of peripheral hearing loss (Pattisapu et al., 2020; Sharma et al., 2020; Torppa & Huotilainen, 2019). When hearing loss is profound, one or two cochlear implants can partially restore hearing capacities. However, auditory perception in cochlear implants (CI) users can remain impaired due to a limited sound frequency discrimination by the implant (Glennon et al., 2020; Lehmann & Paquette, 2015). Indeed, technical constraints of an implant does not allow for a fine-grained decomposition of the auditory signal, compared to the decomposition of a healthy cochlea (McDermott, 2004). As a result, CI users still have difficulties for hearing speech in noise, understanding prosody, and music perception, even though there is a high demand of post-lingually deafened CI users to be able to enjoy music again. These non-verbal auditory perception difficulties are assumed to be related to a pitch perception deficit, resulting from the degraded auditory input. To simulate this pitch perception deficit in normal-hearing (NH) listeners, a signal-processing tool referred to as “vocoder” was used (e.g., Loizou, 2006). Similarly to the CI, vocoders filter the sounds in frequency bands and extract the temporal envelope of the sound (Oxenham, 2008).

Consequently, pitch discrimination thresholds measured in NH participants with vocoded sounds are larger than NH with original sounds, showing similar pitch deficits as CI users (Loizou, 2006; McDermott, 2004; Oxenham, 2008). Here we aim to assess, within a short testing battery, the pattern of non-verbal auditory perception deficits in CI users as well as NH participants with or without vocoded sounds. We thus review below the documented deficits of non-verbal auditory perception in CI users.

1.1. Music perception

Pitch perception has a key role in music processing: pitch discrimination allowing to process frequency differences is essential to perceive the melody and the harmony of music (Marozeau et al., 2013; Oxenham, 2008). In CI users, deficits of music processing affect multiple dimensions. Musical features relating to pitch, such as timbre and harmonicity, are less well perceived by CI users than NH listeners (Hopyan et al., 2016; Petersen et al., 2020; Sharp et al., 2018). Musical time processing was first thought to be relatively preserved in CI (Jiam & Limb, 2019; Spangmose et al., 2019). However, when using complex musical tasks (with pitch variations, not only sequences of beeps), CI users do not perform as well on beat recognition and entrainment (Jiam & Limb, 2019). This result is in link with a reduced mismatch negativity response in CI users with rhythm deviants (Timm et al., 2014).

These deficits contribute to a general depreciation of music and a decreased quality of music listening in CI users (Riley et al., 2018), with a correlation between pitch perception abilities and music appreciation (Zhou et al., 2019). Regarding musical emotions, CI users showed a deficit for recognition compared to NH listeners, especially for sadness (Ambert-Dahan et al., 2015; Hopyan et al., 2016; Shirvani et al., 2014, 2016) or fear stimuli (Paquette et al., 2018). Some studies showed different arousal scores, but similar valence scores in CI users compared to NH listeners (Ambert-Dahan et al., 2015; Lehmann & Paquette, 2015) suggesting that the musical emotion processing deficit would not be a general deficit of music perception. Despite

these limitations, some CI users still enjoy music, and have great demands to enhance this pleasure of music listening (Fuller et al., 2019). Overall, these results suggest that musical emotion recognition is possible in CI, but depends on many features of the implantation as well as of the patient him/herself (Giannantonio et al., 2015). Music processing and appreciation are very variable among CI users. For example, some studies suggest that CI users might have implicit processing of pitch that could help them to enjoy music (Tillmann et al., 2019). In addition, music processing and appreciation might depend on the time of deafness and implantation. For instance, post-lingually deafened and implanted patients complain more about music depreciation compared to early-deafened late implanted patients or prelingually implanted children (Fuller et al., 2019).

Overall, studies of music perception report that music appreciation remains poor in CI users. Even if it depends on previous musical knowledge, music perception, especially related to pitch, could possibly still be improved in CI despite the technical limitations.

1.2. Prosody perception

Even though speech in silent environments is quite well perceived by CI users, the pitch deficit is still limiting their non-verbal auditory perception in speech signals. For intentional prosody, CI users have demonstrated poor perception abilities (Lo et al., 2015; Peng et al., 2008) as well as a deficit for production (Peng et al., 2008). This deficit was associated with difficulties to perceive word and sentence stress (Torppa et al., 2014). For emotional prosody, CI users have deficits when visual cues are unavailable (Pak & Katz, 2019; Paquette et al., 2018). These difficulties are reflected by a change of brain reaction to emotional prosody with a decreased N1-P2 (Deroche et al., 2019). However, it seems that this deficit could be partially compensated with enhanced musical exposure and training (Good et al., 2017).

1.3. Auditory scene perception

Even when speech comprehension in a quiet environment is quite effective, CI users still experience difficulties to hear and understand speech in noise (Bugannim et al., 2019; Choi et al., 2017; Hong & Turner, 2006). Nevertheless, research investigating speech-in-noise perception has shown that CI is helpful for enhancing speech perception in noisy situation (Döge et al., 2017). Even for single-sided deaf patients, the use of CI increased performance in speech perception in noise compared to without CI (Hoth et al., 2016). The benefit was also visible in patients with bilateral hearing loss: bilateral CIs were more efficient compared to patients with bimodal fitting (one CI and one hearing aid) (Choi et al., 2017). However, despite those improvements of hearing in noise with a CI, speech in noise perception or hearing in noise is still very limited (Bugannim et al., 2019; Choi et al., 2017; Hong & Turner, 2006).

As poor verbal working memory and lexical ability can limit speech perception in noise (Kaandorp et al., 2017), tasks without speech were designed to better assess specific difficulties for hearing in noise such as the Music-in-noise task developed in NH participants and musicians (Coffey et al., 2019). Hearing-in-noise perception mostly relies on the ability of the listener to separate auditory sources and focus on the relevant source. Streaming segregation tasks, initially developed for NH users, allow determining how well listeners can discriminate two auditory sources (Van Noorden, 1975). This segregation of auditory sources relies of the ability of the participant to spectrally separate these two sources (Nie & Nelson, 2015). Testing NH participants with vocoded sounds revealed that when the F0 discrimination gets poorer, the segregation between two simultaneous sounds gets very difficult (Gaudrain et al., 2008; Oxenham, 2008), suggesting that this task could be a good index of CI users difficulties. Several studies investigating stream segregation in CI users revealed that CI users have increased perception of one stream and seem to not experience automatic stream segregation (Böckmann-Barthel et al., 2014; Cooper & Roberts, 2007, 2009; Hong & Turner, 2006, 2009). Moreover,

CI users' performance is even worse when background information is fluctuating (Oxenham, 2008). Overall, hearing-in-noise capacities of CI users remain a great challenge that needs to be addressed. Stream segregation tasks appear as an efficient mean to determine hearing in noise capacities in CI users, especially regarding the relevant pitch discrimination capacities (Marozeau et al., 2013).

1.4. Enhancing pitch perception with visual information

Non-verbal auditory perception is still a great challenge for CI users and several studies have been conducted in order to improve this perception. It was suggested that cerebral plasticity could be induced in CI users (pre- and post-lingually) after the implantation (Glennon et al., 2020; Rouger et al., 2012; StreInikov, Marx, et al., 2015). This central plasticity could be one of the major factors explaining the various levels of benefits experienced after implantation. However, the neural mechanisms of this plasticity are still unclear. It appears, however, that cross-modal plasticity in CI users could help to enhance their auditory capacities if correctly directed (Glennon et al., 2020; Strelnikov et al., 2013).

Indeed, to perceive environmental stimuli, multisensory interactions are essential. For instance, the McGurk effect shows that the integration of visual and auditory information is essential for speech perception (McGurk & MacDonald, 1976). Several studies have shown the benefit of multisensory integration for perception rehabilitation (Frassinetti et al., 2005; Grasso et al., 2016; Passamonti et al., 2009). Some findings suggest that multisensory integration is stronger when one of the sensory modalities is deficient. For example, in participants with reduced visual acuity, audiovisual interactions allowed improving their visual detection threshold beyond their visual-only performance, which was not observed in control participants (Caclin et al., 2011). Similarly in participants with a pitch processing deficit, such as congenital amusia, visual stimulations helped them to improve performance in an auditory pitch task (Albouy, Lévêque, et al., 2015). Numerous studies have investigated audiovisual integration in

CI users: First results using speech stimuli and faces showed enhanced auditory capacities in CI users thanks to audiovisual integration (Rouger et al., 2007, 2008; StreInikov, Rouger, et al., 2015; Strelnikov et al., 2009). This audiovisual enhancement is increased in CI users compared to NH participants for all types of speech, even foreign-accented speech (Waddington et al., 2020). One study on non-linguistic speech processing, using voice gender identification, suggests that CI users are influenced more strongly by visual information than are NH participants, even for tasks not directly related to speech comprehension (Barone et al., 2016).

1.5. Rationale for the present testing battery

We designed a new short assessment to test specifically some aspects of non-verbal auditory perception in NH and CI users. This battery aimed to provide a rapid overview of non-verbal auditory capacities in an individual and to be used for further auditory perception research on CI users in particular, thanks to a quick assessment on touch tablet. It measures non-verbal auditory perception capacities with five tests : 1) pitch change detection (based on Albouy, Lévêque, et al., 2015), 2) pitch direction identification (based on Pralus, Lévêque, et al., 2019), 3) short-term memory for pitch (based on Tillmann et al., 2009), 4) stream segregation (based on Grimault et al., 2002) , and 5) emotional prosody (based on Pralus, Fornoni, et al., 2019). Most of these tasks were selected on the basis of previous research investigating pitch processing in congenital amusia. Congenital amusia is a deficit in music perception and production, that is associated to pitch perception (review in Tillmann et al., 2015) and pitch memory deficits (Albouy et al., 2013, 2016; Graves et al., 2019; Tillmann et al., 2009; Tillmann, Lévêque, et al., 2016). These deficits have been traced down to an impaired fronto-temporal network involved in particular in pitch encoding (Albouy et al., 2013, 2019; Albouy, Lévêque et al. 2015). Deficits in pitch change detection (Hyde & Peretz, 2004), pitch direction identification (Loui et al., 2008; Williamson & Stewart, 2010), and pitch short-term memory tasks (Albouy et al., 2013; Graves et al., 2019; Tillmann, Lévêque, et al., 2016) are hallmarks

of the condition of congenital amusia, and we thus selected these three tasks as candidates to uncover the functioning of the combination of the peripheral auditory system and the cortical fronto-temporal network involved in pitch-related auditory cognition. Some research has shown that congenital amusia also affects speech perception, in particular intonation and emotion processing, leading to difficulties for non-verbal auditory cognition (Lima et al., 2016; Lolli et al., 2015; Patel et al., 2008; Pralus, Fornoni et al., 2019; Thompson et al., 2012). Based on prior research with CI users on pitch and music processing, as well as on prosody and auditory scene analysis, we expected to observe deficits in CI users for all five tasks of the battery.

Finally, we also tested whether visual cues could improve pitch processing for CI users as well as NH processing of full and vocoded signals. We used informative and non-informative visual cues in non-verbal pitch perception and memory tasks (pitch change detection, pitch direction identification, and short-term memory) to test the potential usefulness of multisensory integration in these two populations. We hypothesized that CI users would have more difficulties than NH in general but could benefit from the visual cues more strongly than do NH, even than NH in vocoded conditions.

2. Material and methods

2.1. Participants

Ten cochlear implant users and ten normal-hearing participants were recruited (Table 1). All control participants had normal hearing (hearing loss inferior to 20 dB at octave frequencies from 500 to 4000 Hz in both ears, in accordance with BIAP Recommendation 02/1: Audiometric Classification of Hearing Impairments). There were six unilateral CI users with hearing-aids on the contralateral ear, and 4 bilateral CI users, all implanted after post-lingual deafness. We chose CI users with one year or more of implantation to have only CI users in post-implantation chronic phase. All participants (CI users and NH) were selected to have no psychiatric or neuropsychological disorders. Control participants were selected for their good-hearing capacities rather than their age, leading to age differences between the groups (see for example Barone et al., 2016; Butera et al., 2018; Spangmose et al., 2019 for a similar procedure). Study procedures were approved by an ethics committee (randomly selected at the national level, CPP Ile de France VI, ID RCD 2018-A02670-55), NH participants were paid, and CI users were compensated for their participation after having given their written informed consent.

2.2. Testing battery: Material and procedure

The battery was composed of five subtests: Pitch Change Detection (PCD), pitch Direction Change Identification (DCI), pitch Short-Term Memory (STM), Auditory Stream segregation (AS), Emotion recognition (EMO). All tests were implemented to run on an iPad touch tablet, allowing participants to answer by touching large buttons presented on the screen.

For PCD, DCI, and STM tasks, the same stimuli were used, with a roving of frequency across trials. They were synthetic harmonic tones (twelve harmonics), equalized in Root Mean Square (RMS) amplitude, each lasting 500 ms and presented with a within-sequence Inter-Stimulus-Interval (ISI) of 100 ms. Stimuli in AS were based on the same harmonic tones but with a

duration of 100 ms. On half of the trials in PCD, DCI and STM tasks, visual cues were presented. They were composed of white disks with a diameter of 2 cm (100 px) on a black screen. For the experiment with NH participants, we also used vocoded sounds with 16, 8 and 4 channels (see below for details). The experiment took place in a quiet room. The participant was seated in front of the tablet with two loudspeakers (Logitech Z200) at 70 centimeters of distance from the participant's head, with 40 centimeters between each speaker. We set the volume at 55dB SPL for NH participants and adjusted to a comfortable listening level for CI users.

Each participant performed the five subtests in a random order, with each test presenting the stimuli in pseudo-random order, with no more than two repetitions of the same type of stimulus in a row. For NH participants, the order of sound type (normal, vocoded 4 channels, vocoded 8 channels, vocoded 16 channels) was also randomized, all five tests for one sound type were presented in a row, in the same order for each sound types. Before each subtest, participants received an oral explanation of the corresponding task with a written support for CI users and performed a small training to ensure they understood the task. The entire session lasted about 30 minutes for CI users and two hours for NH participants.

Pitch Change Detection (PCD) test

In one trial, participants were presented with a sequence of five isochronous tones, all identical (standard tone) except the fourth tone that could differ in frequency (adapted from Hyde & Peretz, 2004, and Albouy et al., 2015). Standard frequencies were 165, 196, 262 or 392Hz. Deviant frequencies were between 131 and 494 Hz, with changes relative to the standard tone being between 1/16, 1/8, ¼, ½, 1, or 2 tones, either up or down compared to the standard. 64 sequences were constructed with five notes. There were 16 identical trials (four trials per each standard) and 48 different trials (twelve trials per each standard, that is one trial per deviant size, up and down). Non-informative visual cues were presented on half of the trials in addition

to the tones (as in Albouy et al. (2015)). Five circles appeared sequentially from left to right, synchronously with the tones. They were always positioned at the center of the vertical axis on the screen (see Figure 1), and hence were not informative as far as the pitch of the tone was concerned but gave information regarding the onset of the sound. Participants had to determine if the fourth tone of the sequence was the same as or different from the other tones. After the end of the sequence, participants had unlimited time to give their answer by tapping on either the “Same” or the “Different” button. After having given their answer, the next trial was played automatically after an average delay of 1000ms (700-1300ms).

Pitch Direction Change Identification (DCI) test

Participants were presented with two tones at two different frequencies. Fundamental frequencies of the tones were comprised between 123 and 523 Hz. Steps between the two tones could be of 0.5, 1, 1.5, 2, 2.5, 3, or 3.5 tones. 56 sequences were constructed with two different tones, 28 “up” sequences, that is the second tone being higher in pitch than the first one, 28 “down” sequences, that is the second tone being lower in pitch than the first one. Informative visual cues were presented on half of the trials in addition to the tones. Two circles connected by a white bar were appearing consecutively and simultaneously with the onset of each tone. Circles vertical positions were centered on average on the two tones (to be at the center of the screen) and were calculated according the frequency of the corresponding tone, the higher the frequency, the higher the circle on the screen. In contrast to PCD, visual cues were thus fully informative for pitch height (see **Figure 1**), but as they were only present in half of the trials, participants were asked to base their judgements on their auditory perception. These visual cues aimed to reinforce the association between visual height and pitch (Pralus, Lévêque et al., 2019), an effect we plan to exploit in future training experiments. Participants had to determine if the second tone was higher in pitch (Up) or lower (Down) than the first tone. After the end of the second tone, participants had unlimited time to give their answer by tapping on either the

"Up" or the "Down" button. After having given their answer, the next trial was played automatically after an average delay of 1000ms (700-1300ms).

Pitch Short-Term Memory (STM) test

Participants were presented with two melodies of four tones (S1 and S2), with S2 being either identical or different from S1 (adapted from Tillmann et al., 2009; and Hirel et al., 2017). Fundamental frequencies of the tones were comprised between 262 and 440 Hz (corresponding to notes between C4 and A4). 32 melodies were constructed with four tones, each melody thus lasted 2300 ms. In total, there were 16 identical and 16 different trials. For different trials, changes of one tone could occur on the second or third tone. Changes could be of 1.5, 2, 2.5, 3.5, or 4.5 tones, all entailing a change of contour. The delay between the two melodies of a trial was 1000ms. Informative visual cues were presented on half of the trials in addition to the tones, they were presented during S1 and the delay between S1 and S2. Circles were connected by white bars, appearing consecutively and simultaneously with the onset of each tone of the first sequence. Circles' vertical positions were centered on average over the four tones, each vertical circle position was calculated according the frequency of the corresponding tone, the higher the frequency, the higher the circle would be on the screen (see **Figure 1**). Participants had to determine if the second melody was the same or different from the first melody. After the end of the second melody, participants had unlimited time to give their answer by tapping on "Same" or "Different" button. After giving their answer, the next trial was played automatically after an average delay of 1000ms (700-1300ms).

Auditory Stream segregation (AS) test

A sequence was constructed on the model ABA (with A being the standard tone, and B a tone with a varying frequency, both lasting 100 ms). The ISI between A and B was 20ms, the interval between two ABA triplets was 140ms (Grimault et al., 2002; Van Noorden, 1975). Five triplets were repeated for each frequency of B. Fundamental frequency of A was 196 Hz, B was either

196, 247, 294, 440, 659 or 988 Hz (i.e., ranging from 0 to 28 semitones with respect to 196 Hz). The sequence started with a fundamental frequency of B at 440, going down to 196 Hz, then up to 988 Hz, and down again to 196 Hz. This up-and-down pattern was repeated five times, terminating with B having a frequency of 294 Hz. In total, the AS sequence lasted approximately 2.5 minutes. During the sequence, participants had to tell if they hear one or two flux. They gave their answer by tapping on either the "1 flux" or the "2 flux" button. Once one button was selected, it remained selected until the participant changed his/her answer (the selected button stayed highlighted). Participants could respond as many times as they wanted during the sequence.

Emotion recognition (EMO) test

Twenty sentences were taken from Pralus, Fornoni et al. (2019, Experiment 1). These sentences were semantically neutral in French: "J'espère qu'il va m'appeler bientôt" ("I hope he will call me soon"), and "L'avion est presque plein" ("The plane is almost full"). These sentences were uttered with different emotions by male and female actors. For each emotion (joy, neutral, sadness, anger, fear), four sentences were used, half pronounced by a male voice and half by a female voice. Stimuli lasted on average 1470ms (+/- 278ms) and were equalized in RMS amplitude. In each trial, participants listened to a stimulus and were asked to select the recognized emotion from five options (joy, neutral, sadness, anger, fear). After having given their response, they were asked to rate the intensity of the selected emotion from 1 (not intense) to 5 (very intense), except for stimuli judged as neutral (as in Pralus, Fornoni et al. (2019)). They had unlimited time to give their answer. After the intensity rating response, the following stimulus was played automatically after an average delay of 1000ms (700-1300ms).

Vocoded sounds

Three vocoded conditions simulating cochlear implants with different numbers of channels were created, using MatLab R2016a (Mathworks, Inc). For the complete vocoding procedure,

see Massida et al. (2011) and Rouger et al. (2007). All sounds presented in the battery were analyzed through 4, 8, or 16 frequency bands, by using sixth-order IIR elliptical analysis filters. We extracted the temporal envelope by half-wave rectification for each of these frequency bands. The envelope was smoothed with a 500 Hz low-pass third order IIR elliptical filter. We used this extracted envelope to modulate a white noise given by a generator. The obtained signal was filtered with the same filters used previously for the frequency bands. We recombined additively the signals from each frequency band and adjusted the acoustic level obtained to match the original sound level based on RMS.

2.3. Data analysis

We analyzed the data with Bayesian mixed repeated-measures analyses of variance (ANOVA), as implemented in the software JASP (Wagenmakers et al., 2017). In a first set of analyses, we investigated the effect of vocoding in NH participants' data with Sound Type as a within-subject factor, with 4 levels (Non-vocoded, Vocoded with 4, 8, or 16 channels). Other relevant factors depending on the task are detailed below. In the second set of analyses, we compared groups (NH control data for non-vocoded sounds vs. CI), hence the between-participants factor Group was included for all tasks. We report Bayes Factor (BF) as a relative measure of evidence. To interpret the strength of evidence (according to Lee & Wagenmakers, 2014), we considered a BF under three as weak evidence, a BF between three and 10 as positive evidence, a BF between 10 and 100 as strong evidence and a BF higher than 100 as a decisive evidence. BF10 indicates the evidence of H1 (a given model) compared to H0 (the null model), and BFinclusion indicates the evidence of one effect over all models. As no post-hoc tests with correction for multiple comparison have as yet been developed for Bayesian statistics (Wagenmakers et al., 2017, 2018), we used t-tests with Holm-Bonferroni correction for multiple comparisons.

For PCD, DCI, and STM tests, we analyzed the percentage of correct responses with Modality (auditory or audiovisual) as a within-participant factor and the factor Sound Type or Group (as

described above). For PCD (different trials only) and DCI, we ran an additional analysis including the factor of Difficulty (different change sizes). For STM, we also analyzed the percentage of Hits (correct responses for different trials) minus the percentage of False Alarms (incorrect responses for same trials) to correct for potential response bias.

For the AS test, we analyzed first, the total time spent in the perception of one or two flux (thus excluding the time needed to give the first answer), and second, the mean frequency corresponding to changes in the number of flux perceived, with Sound Type or Group as factor.

For the EMO test, we analyzed percentages of correct responses and intensity ratings with Emotion (joy, sadness, anger, fear or neutral) as a within-participant factor and the factor Sound Type or Group. Note that for intensity ratings, the emotion factor had only four levels as neutral stimuli were not rated for intensity. We analyzed only intensity ratings for trials with correctly recognized emotions (as in Pralus, Fornoni et al., 2019). Confusion matrices were calculated based on the percentage of responses given on each type of emotion, compared to the expected emotion.

To uncover the potential links between the five subtests and to understand how they can be used to better characterize the non-verbal auditory perception capacities of the two participant groups (NH participants and CI users), we performed a Principal Component Analysis (PCA) using the FactoMineR package (Lê et al., 2008) on R. We used the percentage of correct responses for auditory trials in PCD, DCI, STM, EMO (recognition), and the time spent in the percept of one flux for AS. Each task and each participant were projected in the multi-dimensional space recovered by the ACP. Correlation factor was also retrieved for each group (CI or NH) for each dimension. We corrected the p-value obtained in this analysis with Holm-Bonferroni correction for multiple comparisons across the five tasks.

In addition, to better understand the relationships between the pitch tasks (PCD, DCI, and STM) with similar auditory stimuli, we performed an additional Bayesian ANOVA on accuracy with Task as a within-participant factor and the factor Group (CI users, NH participants).

3. Results

3.1. PCD test

*Normal-hearing participants and vocoded sounds (**Figure 2A**)*

After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Sound Type (BF10 = 6.47e+7). This model was 3.6 times more likely than the model with the two main effects of Sound Type and Modality (BF10 = 1.8e+7), and 6.5 times more likely than the model with the two main effects and their interaction (BF10 = 1e+7). The other model with the main effect of Modality showed no significant evidence (BF10 = 0.25) (**Table 2**). This was confirmed by a decisive specific effect of Sound Type (BFinclusion = 4.99e+7) only, other specific effects showed no evidence (BFinclusion<0.49). According to t-tests with Holm-Bonferroni correction, original sounds were significantly better recognized than the vocoded sounds with 16, 8, and 4 channels (all pcorr<0.001).

In addition, we analyzed the percentage of correct responses for different trials with the additional factor of Difficulty (six change sizes, see Methods) (**Figure S1A**). After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Sound Type, Difficulty, and the interaction between the two (BF10 = 6.52e+49). It was 8.8 times better than the model with the main effect of Sound Type, Difficulty, Modality, and the interaction between Sound Type and Difficulty (BF10 = 7.37 e+48), and 162 times better than the model with the main effects of Sound Type, Modality, Difficulty, the interaction between Sound Type and Modality, and the interaction between Sound Type and Difficulty (BF10 =

4.02e+47). All of the other models were at least 250 times less likely (BF10<2.6e+47). This was confirmed by a decisive specific effect of Difficulty (BFinclusion = 1.3e+14), Sound Type (BFinclusion = 1.3e+14), and the interaction between Sound Type and Difficulty (BFinclusion = 9.8e+4). Other specific effects showed no significant evidence (BFinclusion<0.04). According to t-tests with Holm-Bonferroni correction, original sounds were significantly better recognized than the vocoded sounds with 16, 8, and 4 channels (all pcorr<0.001). Over all types of sounds (original or vocoded), trials of difficulty of 1/16 and 1/8 tone lead to poorer performance than all other trials (all pcorr<0.017), trials of difficulty of 1/4 tone were less well performed than trials of difficulties of 1 and 2 tones (both pcorr<0.009), trials of difficulties of ½ and 1 tone were less well performed than trials of difficulty of 2 tones (both pcorr<0.019). For original sounds, trials of difficulty of 1/16 were less well performed than that with difficulties of ¼ and 2 (both pcorr<0.038). For 16-channels vocoded sounds, trials of difficulties of 1/16 and 1/8 were less well performed than trials of difficulties of ½, 1 and 2 (all pcorr<0.042), trials of difficulty of 1/16 were less well performed than that with difficulty of ¼ (pcorr<0.001). For 8-channels vocoded sounds, trials of difficulties of 1/16, 1/8 and 1/4 were less well performed than that of difficulties of 1 and 2 (all pcorr<0.01), trials of difficulties of 1/16 and 1/8 were less well performed than that of difficulty of 1/2 (both pcorr<0.042). For 4-channels vocoded sounds, trials of difficulties of 1/16, 1/8 and 1/4 were less well performed than that of difficulties of 2 (all pcorr<0.001), trials of difficulty of 1/8 were less well performed than that of difficulties of ½ and 1 (both pcorr<0.025). Thus overall, when less channels were used for the vocoded sounds, discrimination got harder even with large physical differences.

*Cochlear implant listeners compared to normal-hearing participants (**Figure 3A**)*

After comparison to the null model, the only model showing strong evidence was the one with the main effect of Group, Modality, and the interaction between the two (BF10 = 24.9). All other models showed no significant evidence (BF10 < 1) (**Table 3**). This was confirmed by a

positive specific effect of Modality (BFinclusion = 8.2), and strong specific effects of Group (BFinclusion = 13.1) and of the interaction between the two (BFinclusion = 35.1). According to t-tests with Holm-Bonferroni correction, CI had significantly lower performance than NH for audio stimuli (pcorr=0.03). CI were significantly better with audiovisual stimuli compared to audio stimuli (pcorr=0.01). NH tended to be slightly better with audiovisual stimuli compared to audio stimuli (pcorr=0.08). Interestingly, performance between CI and NH for audiovisual stimuli did not differ.

In addition, we analyzed the percentage of correct responses for different trials with the additional factor of Difficulty (six change sizes) (**Figure S2A**). After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Group, Difficulty, and the interaction between the two (BF10 = 1.24e+16). It was 1.6 times better than the model with the main effect of Difficulty (BF10 = 7.68e+15), 3.4 times better than the model with the main effects of Group and Difficulty (BF10 = 3.61e+15), and 5.8 times better than the model with the main effect of Group, Difficulty, Modality and the interaction between Group and Difficulty (BF10 = 2.14e+15). All other models were at least 10 times less likely (BF10<1.2e+15). This was confirmed by a decisive specific effect of Difficulty (BFinclusion = ∞) and a weak specific effect of the interaction between Group and Difficulty (BFinclusion = 2.5). Other specific effects showed no significant evidence (BFinclusion<0.8). According to t-tests with Holm-Bonferroni correction, over the two groups, trials of difficulty of 1/16 tone were less well categorized than all other trials (all pcorr<0.004), trials of difficulty of 1/8 tone were less well categorized than trials of difficulties of 1 and 2 tones (both pcorr<0.012). In CI users, we found the same pattern of responses: trials of difficulties 1/16 and 1/8 were less well categorized than other difficulties (all pcorr<0.025). In NH participants, only trials of difficulty of 1/16 tone were less well categorized than trials of difficulty of ¼, ½, 1 and 2 tones (all pcorr<0.026).

3.2. DCI test

*Normal-hearing participants and vocoded sounds (**Figure 2A**)*

After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Modality (BF10 = 2.78e+6). It was 8.9 times better than the model with the main effect of Modality and Sound Type (BF10 = 3.12e+5), and 43 times better than the model with the two main effects of Sound Type and Modality and their interaction (BF10 = 6.42e+4). The model with the main effect of Sound Type only showed no significant evidence (BF10 = 0.1) (**Table 2**). This was confirmed by a decisive specific effect of Modality (BFinclusion = 1.92e+6) only. Other specific effects showed no significant evidence (BFinclusion<0.09).

In addition, we analyzed the percentage of correct responses with the additional factor of Difficulty (seven change sizes, see Methods) (**Figure S1B**). After comparison to the null model, the best model showing decisive evidence was the one with the main effects of Modality, Difficulty, and the interaction between the two (BF10 = 6.36e+23). It was 41.3 times better than the model with the main effects of Modality, Sound Type, Difficulty, and the interaction between Modality and Difficulty (BF10 = 1.54e+22), and 468 times better than the model with the main effects of Modality, Sound Type, Difficulty, and the interaction between Sound Type and Modality, and between Modality and Difficulty (BF10 = 1.36e+21). All of the other models were at least 1870 times less likely (BF10<3.4e+20). This was confirmed by decisive specific effects of Modality (BFinclusion = 4.6e+14), Difficulty (BFinclusion = 3.8e+4), and the interaction between Modality and Difficulty (BFinclusion = 4020). Other specific effects showed no significant evidence (BFinclusion<0.01). According to t-tests with Holm-Bonferroni correction, trials of difficulties of 0.5 and 1 tone were less well categorized than trials of difficulties of 2.5, 3 and 3.5 tones (all pcorr<0.05). Audiovisual trials were specifically better categorized than auditory trials for difficulty levels of 0.5, 1, 1.5 and 3 tones (all pcorr<0.001).

*Cochlear implant listeners compared to normal-hearing participants (**Figure 3A**)*

After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Modality (BF10 = 4178.2). It was 1.5 times better than the model with the main effect of Modality and Group (BF10 = 2708.5), and 2 times better than the model with the two main effects of Group and Modality and their interaction (BF10 = 2113.5). The model with the main effect of Group only showed no significant evidence (BF10 = 0.5) (**Table 3**). This was confirmed by a decisive specific effect of Modality (BFinclusion = 3952.3) only. Other specific effects showed no significant evidence (BFinclusion<1.2). As the three first models were very close, we decided to report the post-hoc comparisons for the interaction. According to t-tests with Holm-Bonferroni correction, both groups had better scores with audiovisual trials compared to auditory trials (pcorr<0.001 for CI users and pcorr=0.013 for NH). CI users were significantly better in audiovisual trials compared to NH participants with auditory trials (pcorr=0.007), but not when comparing their performance in auditory trials to NH's performance in auditory trials. Note that on average across audiovisual and auditory trials, CI users were thus performing better than NH participants.

In addition, we analyzed the percentage of correct responses with the factor of Difficulty (seven change sizes, see Methods) (**Figure S2B**). After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Modality and Difficulty and the interaction between the two (BF10 = 2.28e+13). It was 1.52 times better than the model with the main effect of Modality (BF10 = 1.5e+13), 1.8 times better than the model with the main effects of Modality, Group and Difficulty and the interaction between Group and Modality and between Modality and Difficulty (BF10 = 1.26e+13), and 1.9 times better than the model with the main effects of Modality, Group and Difficulty and the interaction between Modality and Difficulty (BF10 = 1.19e+13). All of the other models were at least 2.7 times less likely (BF10<8.4e+12). This was confirmed by a decisive specific effect of Modality (BFinclusion =

1.87e+13) and a small positive effect of the interaction between Modality and Difficulty (BFinclusion = 3). Other specific effects showed no significant evidence (BFinclusion<0.77). According to t-tests with Holm-Bonferroni correction, trials of difficulty of 0.5 tone had a tendency to be less well categorized than trials of difficulties of 2.5 tones (pcorr=0.14). Audiovisual trials were specifically better categorized than auditory trials for difficulty levels of 0.5, 1 and 1.5 (all pcorr<0.013).

3.3. STM test

*Normal-hearing participants and vocoded sounds (**Figure 2A**)*

None of the tested models explained the data (percentage of correct responses) better than the null model (BF10 < 1) (**Table 2**). This was confirmed by no significant specific effects (BFinclusion<0.3).

In addition, we analyzed the percentage of Hits minus the percentage of false alarm in NH participants with all sound types, in audio or audiovisual conditions (**Figure 2B**). None of the tested models explained the data better than the null model (BF10 <1.7). This was confirmed by no significant specific effects (BFinclusion<1.18).

*Cochlear implant listeners compared to normal-hearing participants (**Figure 3A**)*

None of the tested models explained the data (percentage of correct responses) better than the null model (BF10 < 1) (**Table 3**). This was confirmed by no significant specific effects (BFinclusion<0.7).

For the analysis of the percentage of Hits minus the percentage of False Alarms (**Figure 3B**), after comparison to the null model, the best model showing positive evidence was the one with the main effect of Modality (BF10 = 7.1). This model was 1.7 times more likely than the model with the two main effects of Modality and Group (BF10 = 4.3). The other models showed no significant evidence (BF10<2). This was confirmed by a positive evidence of Modality only (BFinclusion = 5.8), other specific effects showed no evidence (BFinclusion<0.6). Performance

in the audiovisual trials was better than performance in the auditory trials in both groups. NH participants seemed to have generally more hits minus false alarms (i.e. better performance) than CI users.

3.4. AS test

Normal-hearing participants and vocoded sounds

*Total time spent in one or two flux percepts (**Figure 2C**).* After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Percept (one or two flux) and Sound Type, and the interaction between the two (BF10 = 3.28e+48). This model was 7.2 times more likely than the model with the main effect of Percept (BF10 = 4.3e+47), and 94 times more likely than the model with the two main effects of Percept and Sound Type (BF10 = 3.48e+46). The model with the main effect of Sound Type showed no evidence (BF10 = 0.07) (**Table 2**). This was confirmed by a decisive specific effect of Percept (BFinclusion = 9.8e+13), a strong specific effect on the interaction between Perception and Sound Type (BFinclusion = 28) and a positive specific effect of Sound Type (BFinclusion = 5.1). According to t-tests with Holm-Bonferroni correction, less time was spent in the perception of one flux compared to two flux for all types of sounds (all pcorr <0.001). There was a tendency to spend less time in the percept of one flux with 4-channel vocoded sounds compared to original sounds (pcorr=0.069), and a tendency to spend more time in the percept of two flux with 4-channel vocoded sounds compared to 16-channel vocoded sounds (pcorr=0.069).

*Mean frequency at the change of perception (**Figure 2D**).* After comparison to the null model, the model with the main effect of Sound Type showed a strong evidence (BF10 = 12.85) (**Table 2**). According to t-tests with Holm-Bonferroni correction, there were a tendency for a higher mean frequency with 8-channel vocoded sounds compared to 16-channel vocoded sounds (pcorr=0.1).

Cochlear implant listeners compared to normal-hearing participants

*Total time spent in one or two flux percepts (**Figure 3C**).* After comparison to the null model, the best model showing decisive evidence was the one with the main effect of Percept (one or two flux) and Group, and the interaction between the two (BF10 = 6.4e+16). This model was 9.7 times more likely than the model with the main effect of Percept (BF10 = 6.6e+15), and 27.8 times more likely than the model with the two main effects of Percept and Group (BF10 = 2.4e+15). The model with only the main effect of Group showed no evidence (BF10 = 0.3) (**Table 3**). This was confirmed by a decisive specific effect of Percept (BFinclusion = ∞), a strong specific effect on the interaction between Percept and Group (BFinclusion = 28.8) and a positive specific effect of Group (BFinclusion = 6.8). According to t-tests with Holm-Bonferroni correction, both groups spent less time in the perception of one flux compared to two flux (both pcorr <0.001). CI users spent significantly less time in the perception of two flux compared to NH controls (pcorr=0.034), and more time in the perception of one flux compared to NH controls (pcorr=0.024).

*Mean frequency at the change of perception (**Figure 3D**).* After comparison to the null model, the model with the main effect of Group showed a positive evidence (BF10 = 5.1) (**Table 3**). The frequency difference between A and B at which switches between percepts occurred was higher for the CI group (111.6 Hz) than NH group (49.8 Hz).

3.5. EMO test

Normal-hearing participants and vocoded sounds

*Emotion categorization (**Figure 2E**).* After comparison to the null model, the best model showing a decisive evidence was the one with the two main effects of Sound Type and Emotion and their interaction (BF10 = 5.68e+19). This model was 7.8 times more likely than the model with the two main effects of Sound Type and Emotion (BF10 = 7.3e+18) and 14200 times more likely than the model with the main effect of Sound Type (BF10 = 4e+15). The model with the main effect of Emotion showed strong evidence (BF10 =19.9) (**Table 2**). This was confirmed

591 by a decisive specific effect of Emotion (BFinclusion = 1.06e+4) and Sound Type (BFinclusion
592 = 3.3e+14), and a strong specific effect of the interaction between Emotion and Sound Type
593 (BFinclusion = 31.1). According to t-tests with Holm-Bonferroni correction, Anger was
594 significantly better recognized than Joy and Neutrality (pcorr=0.019 and 0.008 respectively).
595 Original sounds were significantly better recognized than the vocoded sounds with 16, 8 and 4
596 channels (all pcorr<0.003). Vocoded sounds with 16 channels were significantly better
597 recognized than the ones with 8 and 4 channels (both pcorr<0.001) and vocoded sounds with 8
598 channels were better recognized than the ones with 4 channels (pcorr=0.016). Normal and
599 vocoded sounds with 16 channels were significantly better recognized than vocoded sounds
600 with 4 channels for Joy, Sadness and Neutrality (all pcorr<0.003). Original sounds and 16
601 channels-vocoded sounds were significantly better recognized than 8 channels-vocoded sounds
602 for Joy (pcorr=0.023). Original sounds were significantly better recognized than 8 channels-
603 vocoded sounds for Neutrality (pcorr=0.011).

604 Confusion matrices (**Table 4**) showed that with vocoded sounds, fear was often confused with
605 anger, which was never the case with original sounds: indeed with original sounds fear was
606 more confused with sadness. Moreover, with vocoded sounds, sadness was often confused with
607 neutrality, whereas this confusion was not present with original sounds.

608 *Intensity ratings (**Figure 2F**).* After comparison to the null model, the best model showing a
609 decisive evidence was the one with the two main effects of Sound Type and Emotion and their
610 interaction (BF10 = 2.8e+5). This model was 1202 times more likely than the model with the
611 two main effects of Sound Type and Emotion (BF10 = 233.6) and 7777 times more likely than
612 the model with the main effect of Sound Type (BF10 = 36.2). The model with the main effect
613 of Emotion showed small positive evidence (BF10 =3.7) (**Table 2**). This was confirmed by a
614 decisive specific effect of Emotion (BFinclusion = 4973), Sound Type (BFinclusion = 3.4e+4),
615 and the interaction between Emotion and Sound Type (BFinclusion = 4038). According to t-

tests with Holm-Bonferroni correction, Anger was significantly rated more intense than Fear and Sadness (pcorr=0.04 and 0.004 respectively). Joy was rated more intense than Sadness (pcorr=0.005). 16-channels vocoded sounds tended to be rated higher than 4-channels vocoded sounds (pcorr=0.078). Vocoded sounds with 16 channels were rated significantly higher than vocoded sounds with 4 channels for Joy and Sadness (both pcorr<0.001).

Cochlear implant listeners compared to normal-hearing participants

*Emotion categorization (**Figure 3E**).* After comparison to the null model, the best model showing a decisive evidence was the one with the two main effects of Group and Emotion and their interaction (BF10 = 4107.7). This model was 7 times more likely than the model with the two main effects of Group and Emotion (BF10 = 588.8) and 60 times more likely than the model with the main effect of Emotion (BF10 = 68). The model with the main effect of Group showed positive evidence (BF10 = 8.7) (**Table 3**). This was confirmed by a decisive specific effect of Emotion (BFinclusion = 325.9), a strong specific effect of Group (BFinclusion = 45.5) and of the interaction between Emotion and Group (BFinclusion = 24.7). According to t-tests with Holm-Bonferroni correction, Fear was significantly less recognized than Anger and Neutrality (all pcorr<0.001). CI had lower recognition scores compared to NH for Joy (t(8)=4.3 pcorr=0.004) and for Sadness (t(8)=3.5 pcorr=0.038) (other pcorr > 0.8).

Confusion matrices (**Table 5**) showed that in CI users, joy was often confused with sadness, which was never the case in control participants. Moreover, in CI users, sadness was often confused with neutrality, this confusion was not observed for NH participants.

*Intensity ratings (**Figure 3F**).* After comparison to the null model, all models showed no significant evidence (BF10 < 1.1) (**Table 3**). This was confirmed by no significant specific effects (BFinclusion<0.8).

3.6. Relationships between the tasks

*Principal Component Analysis (**Figure 4**)*

The PCA revealed that three dimensions explained 84% of the variance. In particular, performance in the three tasks PCD, EMO and AS correlated with the first dimension (all pcorr<0.025), as well as the Group variable (p=0.0098), with NH participants performing better at these tasks than CI users. The Group variable (p=0.026) as well as performance in the tasks DCI and STM (both pcorr<0.015) correlated with the second dimension. However, on this dimension, CI users performed better than NH participants. The STM task also correlated with the third dimension (pcorr=0.03). These results suggest that the tasks of PCD, EMO and AS provided similar information regarding non-verbal auditory perception and allow for a distinction between the group of NH participants and CI users. This first dimension reflects a pitch discrimination deficit in CI users compared to NH participants. Similarly, STM and DCI also seem to reveal a different pitch perception in these two groups. This second dimension seems to be more related to contour information as this was relevant for these two tasks. Yet the third dimension of the PCA suggests that STM alone could bring supplementary information (for both groups). Indeed, for STM, the memory load is the biggest compared to all tasks, but this does not seem to affect the groups differently. Overall, this analysis reveals a deficit of CI users for pitch discrimination task, but a better analysis of contour of CI users compared to NH participants. This finding suggests that CI users were cognitively intact but had specific difficulties regarding the pitch discrimination.

Cochlear implant listeners compared to normal-hearing participants in pitch tasks (PCD, DCI and STM)

After comparison to the null model, the best model showing a positive evidence was the one with the two main effects of Group and Task and their interaction (BF10 = 4.08). The other models showed no significant evidence (BF10<2.6). This was confirmed by positive specific effects of Task (BFinclusion=3.59) and the interaction between Group and Task

(BFinclusion=3.03), Group showed no significant specific effect (BFinclusion=1.1). According to t-tests with Holm-Bonferroni correction, the PCD showed greater recognition scores than DCI (pcorr=0.046) and STM (pcorr=0.05). In the control group, the PCD task showed greater recognition scores than the DCI (pcorr=0.004). This was not the case in the CI users, reflecting more homogenous results across tasks in this group, with higher scores in DCI but also smaller scores in PCD compared to NH participants.

4. Discussion

Overall, our study demonstrated that rapid assessment of non-verbal auditory perception can be done in NH participants and CI users. CI users showed prominent deficits in three out of the five battery tasks compared to NH participants: Pitch Change Detection, Auditory Stream segregation, Emotion (prosody) recognition. In contrast, they reached similar performance levels for the pitch Direction Change Identification task, as well as the pitch sequence Short-Term Memory task in comparison to NH participants. This pattern of perceptual deficits of CI users was mostly mimicked in NH participants with vocoded sounds, with a deficit partly correlated to the number of channels of the vocoder used. In addition, both groups benefited from visual cues in pitch tasks, but these effects were particularly prominent in the CI group.

4.1. Patterns of non-verbal auditory perception deficits in CI users and in NH participants hearing vocoded sounds

In the PCD test, which was the most basic task related to pitch perception in the battery, CI users demonstrated a deficit compared to NH participants. CI users' performance level was comparable to performance level of NH participants listening to vocoded sounds (Figures 2A, 3A). These deficits were more pronounced for difficult trials: the smaller the size change was, the bigger were the deficits of CI users and NH participants with vocoded sounds. These results were expected as CI and vocoded sounds only give a partial and degraded information about the pitch of the sound (Glennon et al., 2020; Lehmann & Paquette, 2015). As we used roving pitches in the PCD task, we prevented frequency-related training, and showed that the deficit was not specific to one frequency in particular.

Interestingly, in the DCI task, participants did not need to recognize the pitch per se, but just needed to distinguish a pitch difference and to infer a direction on this difference. CI users showed no deficit on this task compared to NH participants who listened either to the original sounds or the vocoded sounds. This could be linked to the fact that, compared to the PCD task,

participants were not asked to detect a pitch change, but rather to make a decision on a relation between two tones, recognize the contour. Previous work demonstrated that CI users had remaining implicit pitch processing (Tillmann et al., 2019), which could have helped in this task. Moreover, in the DCI task, the pitch change sizes were bigger than in the PCD task. However, this task was not easier than the PCD, as reflected by the combined analysis of PCD, DCI and STM. Indeed, for NH participants, it was more difficult to correctly identify the direction of a pitch change (DCI), than detect a pitch change (PCD).

As pitch is essential for music perception, and CI users often complain about their poor musical appreciation (Paquette et al., 2018; Riley et al., 2018; Zhou et al., 2019), we used a short music-like task to test their perception of melody in the context of short-term memory (STM) task. Interestingly, CI users did not show a deficit on this task and performed similarly as did NH participants with original sounds. Note that NH participants did not show a deficit on this task with vocoded sounds either. Even though hearing loss can decrease cognitive abilities (Amieva & Ouvrard, 2020), various studies have demonstrated that compensating this hearing loss, in particular with CI, can improve cognitive performance (Amieva & Ouvrard, 2020; Claes et al., 2018; Mosnier et al., 2015; Völter et al., 2018). The observed data pattern here suggests that the short-term memory abilities of CI users were sufficient to perform well at this memory task, even though the difference of melodies was based on a change on the pitch dimension, that is an acoustic dimension difficult to process for them. Moreover, as the difference included a change of contour, this could have helped CI users to detect this difference, similarly as in the DCI task. Indeed, previous work on melodic contour identification (MCI task) revealed that CI users are able to correctly identify melodic contour (Galvin et al., 2007; Luo et al., 2014; Wright & Uchanski, 2012). Their recognition scores can be similar to NH participants when the number of semitones between notes is large enough (Galvin et al., 2007), and improve with the number of years of musical experience (Galvin et al., 2009). Overall, it appears that when the task is

not purely pitch-perception centered, CI users can perform as well as NH participants. This could explain why, despite their poor pitch discrimination abilities, CI users are still able to listen to and enjoy music (Fuller et al., 2019; Wright & Uchanski, 2012). It might be argued that the patients were not impaired severely enough so that they still succeeded. However, the here tested CI users experienced difficulties with the PCD task and the two tasks AS and EMO, which are both more related to their everyday life perception.

In agreement with previous reports (Böckmann-Barthel et al., 2014; Cooper & Roberts, 2007, 2009; Hong & Turner, 2006, 2009), our AS task detected the deficit of streaming segregation in CI users. Indeed, CI users spent more time in the perception of one flux, meaning that their segregation was not performed as efficiently as NH participants. This was confirmed by an increase of the frequency at change of percept. Interestingly, these results were also found in NH participants with vocoded sounds, but only for the 4-channels vocoded sounds, i.e., the most degraded sounds. This reflects the deficit of CI users in everyday life to segregate two auditory sources (Cooper & Roberts, 2007, 2009; Oxenham, 2008) and more generally to hear signals in noise. However, it appears that everyone, even CI users spent a long part of their life in noisy situations, mostly trying to understand sounds in background noise (Busch et al., 2017). Even if CI users have big difficulties to hear in noise, they are still able to manage this task in some cases (Döge et al., 2017; Hoth et al., 2016). To reflect this deficit, streaming segregation tasks have often been used as they reflect a way of understanding hearing in noise apart from the speech comprehension per se. Some studies suggest that these tasks only show a need of more time to make a decision and subjective uncertainty in CI users compared to NH participants (Böckmann-Barthel et al., 2014). However, to account for this, we decided to make several back and forth presentations of the same pitch differences between the A and B sounds of the ABA triplets (Grimault et al., 2002), and to measure perception over the entire sequence. Here, we can assume that the subjective uncertainty would be compensated over the time of the

sequence. Recent studies also showed that this decreased streaming in CI users was not only due to an increased decision time as it depends on electrode separation but also on tone repetition time (Paredes-Gallardo et al., 2018b, 2018a; Saki et al., 2019), similarly as in NH participants (Van Noorden, 1975). Moreover, even if this kind of perceptual task is rather simple, performance correlates with speech perception in noise (Hong & Turner, 2006; Saki et al., 2019). These findings taken together suggest that streaming segregation is a simple and rapid task easy to evaluate the capacities of hearing in noise in CI users, independently of their phonological skills.

Finally, as pitch is important for non-verbal auditory cues in speech, we also evaluated the perception of emotional prosody in our participants. Using a simple paradigm of emotion categorization of short sentences, we demonstrated a deficit of emotion recognition in CI users compared to NH participants. This deficit was particularly pronounced for joyful and sad sentences, which CI users tended to confound more frequently than did NH participants. Interestingly, the same pattern of deficit was found in NH participants with vocoded sentences, with a deficit related to the number of vocoded channels, but also a specific deficit for neutrality. This demonstrates efficiently the poor perception of emotional prosody in CI users already documented before (Deroche et al., 2019; Everhardt et al., 2020; Pak & Katz, 2019; Paquette et al., 2018), and allows for a better characterization of this deficit depending on the emotion. These results can be directly linked to previous findings reporting a reduced N1-P2 component in CI users for sadness, but also a reduced late brain response in CI users, generally associated to the distinction between joy and sadness in NH participants (Deroche et al., 2019). Moreover, here the double paradigm using both emotion categorization and intensity ratings allowed us to show that perception of emotional prosody is not fully disrupted in CI users. This data pattern replicates our previous finding of a study on a different population with pitch processing deficit: congenital amusia (Pralus et al., 2019). Indeed, individuals with congenital amusia

773 demonstrated deficit of emotional prosody recognition with short vowels but no deficit for
774 intensity ratings (Pralus et al., 2019). Similarly, intensity ratings of CI users and NH participants
775 did not differ. This result reflects the potential capacities of CI users to correctly perceive
776 emotional prosody on an implicit level. Indeed, to rate the intensity of an emotion, no conscious
777 representation of a given stimulus was required. Intensity ratings are more linked to lower
778 activation levels of the representation of this stimulus. Thus the results of intensity ratings in
779 CI users reflect a potential implicit preservation of pitch processing in the case of emotional
780 prosody (Tillmann et al., 2019), already demonstrated in another deficit of pitch perception:
781 congenital amusia (Frühholz et al., 2012; Pralus et al., 2019; Stewart, 2011; Tillmann, Lalitte,
782 et al., 2016). Interestingly, the pattern of results was different for NH participants with vocoded
783 sentences. Indeed, these participants showed decreased intensity ratings for joy and sadness
784 with 4-channels vocoded sentences. This could be explained by the fact that NH participants
785 are not used to hear degraded speech noise and use preferentially explicit strategies to detect
786 and judge intensity of an emotion. As vocoded sentences could seem a little dehumanized for
787 NH participants, they could have assumed that emotions were less sincere and intense,
788 especially for joy which is generally very intense and for sadness which can be confounded
789 with neutral. This is also in line with a previous work that demonstrated the limitations of
790 vocoder strategy to simulate the CI in NH participants, revealing that vocoded sounds did not
791 give the same results as CI regarding music appreciation (Wright & Uchanski, 2012).

792 ***4.2. Processes involved in non-verbal auditory perception: pitch discrimination, pitch***
793 ***contour, and memory***

794 The analysis of each test revealed that CI users' performance was impaired for PCD, EMO and
795 AS tasks, but did not differ from NH participants' performance for DCI and STM tasks. To
796 better characterize this performance pattern, we ran an ACP on the five tasks (auditory trials
797 only). This analysis confirmed a specific deficit of the CI users compared to the NH participants

for the tasks of PCD, EMO and AS, and this deficit correlated with the first dimension of the analysis. This first dimension shows that CI users experience difficulties with pitch discrimination tasks, and this deficit can influence the recognition of emotion but also the capacity to segregate auditory sources. Interestingly, the second dimension of the ACP solution revealed that CI users were better than NH participants in some aspects of the DCI and STM tasks. These two tasks can be performed using the contour between two or more notes. The results revealed that CI users can use this contour information as well (or even better) than NH participants, as already demonstrated earlier with more musical tasks (Galvin et al., 2007; Luo et al., 2014; Wright & Uchanski, 2012). Moreover, the third dimension of the ACP correlated only with the STM task, revealing its specific testing aspect regarding pitch memory. The STM task involves more cognitive load with memory for pitch. However, on this third dimension, there were no differences between our two groups. Overall, these results show that the five tests can partially depict the same pattern of pitch perception deficit, but they also distinguish different part of this perception. The three pitch tasks (PCD, DCI, STM) can be used to summarize the pitch deficits of CI users compared to NH participants with respect to the processes of discrimination, contour, and memory. As we used the same kind of stimuli for the three tasks, a group analysis on these three tasks was performed. It revealed that CI users had specific deficit with PCD, but this deficit did not affect as strongly performance in the DCI task. This resulted in the overall results of CI users' performance being similar in the three tasks, whereas the NH participants were specifically better in PCD. These two supplementary analysis thus suggest that while CI users have impaired pitch detection performance, which can be directly linked to the poor pitch processing with the CI, they performed as well as (or even outperform) NH participants when the tasks involved more cognitive processing, such as contour or memory processing. The test combination of this battery is thus also interesting to detect specifically a pitch processing deficit in CI users and to separate this deficit from more

general differences in cognitive abilities related to the hearing loss. Here in this group of CI users without neuropsychological deficits, it reveals that despite their pitch discrimination deficits, CI users might rely on pitch contour information processing (for large intervals) more than NH listeners.

Overall, the present battery allows characterizing rapidly (30 minutes) and easily (using a touch tablet) several facets of non-verbal auditory perception in CI users and NH participants. The battery includes basic pitch detection and categorization tasks, but also tasks closer to listening experience of everyday life, such as emotional prosody perception and the challenge of sound source segregation.

4.3. Benefit of audiovisual cues for non-verbal auditory tasks

In the present test battery, we also investigated a potential enhancement of pitch perception by additional visual stimulation. In the PCD task, we used non-information visual cues, which provided only temporal cues about the tones (and in particular, their onset and duration). Interestingly, these visual cues were sufficient to significantly enhance CI users' detection score, compared to their detection score with auditory trials. This pattern of results thus extends the previously reported visual cue benefit from participants with congenital amusia (Albouy, Lévêque, et al., 2015) to CI users. While CI users showed a deficit of pitch change detection with auditory trials, they reached similar performance with audiovisual cues than did NH participants. We did not find this effect in NH participants with vocoded sounds, demonstrating that CI users could benefit more efficiently from visual cues as they might be more used to rely on them based on their everyday life experience (Rouger et al., 2007). Even if the visual cues were not informative regarding the pitch of the tones, the visual information with its adequate timing could have boosted participants' dynamic attending toward the onsets of the tones as the tones were presented in a regular, isochronous sequence in the PCD task (Fiveash et al., 2020; Jones, 1976; Kuroda et al., 2017).

In the STM task, the visual cues were informative for the pitch dimension, but they were present only during the first melody and the retention delay. Hence, they did not inform about the correct answer (which was the case in the DCI task). For the STM task, we observed a general improvement of scores with audiovisual trials compared to auditory trials in both populations. This result demonstrates that informative visual cues can help participants to enhance auditory perception and encoding, leading to better memory recognition performance, even though these cues were not present during the presentation of S2. In the DCI task, visual cues were fully informative: they were sufficient to do the task, without listening to the audio. Results revealed an improvement with audiovisual trials compared to auditory trials in both groups, and for NH participants also with vocoded sounds. The visual benefit was even more pronounced in CI users who even outperformed NH participants with auditory trials. As also suggested by the PCD task data, CI users seem to better integrate multisensory information and benefit more strongly from this integration than did the NH participants (Barone et al., 2016; Rouger et al., 2007; StreInikov, Rouger, et al., 2015). Indeed, the scores of NH participants with audiovisual trials were not at ceiling, suggesting that these participants might still be relating on auditory perception to do the task (as they were requested to do) and made less use of the visual information.

In conclusion, results from these three tasks showed that visual cues, informative or not, can boost the performance of participants in pitch perception tasks. Moreover, this boost effect can be particularly efficient in CI users.

Over the past few years, many training and rehabilitation strategies have been developed to enhance pitch perception abilities in CI users. For instance, some training focused on auditory musical training (Barlow et al., 2016; Cheng et al., 2018; Firestone et al., 2020; C. D. Fuller et al., 2018; Galvin et al., 2009; Good et al., 2017; Lerousseau et al., 2020; Lo et al., 2015; Patel, 2014). The findings of these training programs have demonstrated some improvement of

intentional (Lo et al., 2015) or emotional prosody recognition in CI users (Good et al., 2017). Other studies proposed training strategies based on multi-sensorial integration, previously demonstrated to be enhanced in CI users (Rouger et al., 2007). For example, audio-motor integration (Chari et al., 2020) using electrotactile stimulation showed some minor improvement in speech perception (Huang et al., 2017). Based on previous results in CI users (Innes-Brown et al., 2011; Rouger et al., 2007, 2008; StreInikov, Rouger, et al., 2015) and on the present results of our audiovisual tasks, it seems that audiovisual training could be a good strategy to enhance auditory perception in CI users (Sato et al., 2019). The present results are in favor of a training strategy based on informative cues, as demonstrated in DCI and STM tasks. These informational visual cues could benefit to pitch perception, especially when the difficulty is increased. Nevertheless, a training strategy based on non-informative cues could be enough to allow for multisensory integration and to enhance pitch perception. In conclusion, these perception data of the battery provide some interesting insights for further training strategies in CI users, as well as for other populations with pitch perception deficits (e.g., individuals with congenital amusia).

5. Conclusion

Our present findings suggest that our battery can be used to characterize non-verbal auditory perception in participants with hearing difficulties (here CI users), as well as in NH participants with vocoded sounds. This assessment allows for a rapid (30 minutes) detection of pitch difficulties with auditory non-verbal sounds, as well as emotional prosody processing and stream segregation capacities. Moreover, the three pitch tasks using visual cues allowed us to better characterize multisensory integration in NH participants and CI users. In particular, non-informative cues used in the pitch detection task enhanced pitch perception both in CI users and in NH participants with vocoded sounds. These results could be a good starting point to devise a new audiovisual training procedure for participants with pitch perception difficulties (Fuller

et al., 2018), notably the types of audiovisual cues to use (informative and non-informative). The present battery can then be used to monitor the enhancement of pitch perception abilities in participants (i.e., pre/post-training assessment; Pralus et al., in progress).

6. Acknowledgements

We would like to kindly thank Lesly Fornoni for her help to recruit the NH participants, and Julie Gatel for her help to recruit the CI users.

A.P. designed and performed experiments, analyzed data and wrote the paper; R.H. helped for the recruitment, the inclusion and the experiments with cochlear implant users. F.C. performed experiments with the NH participants. P.E.A realized the technical application of the experiment on touch tablet. A.M provided technical support for the experiment and critical revision. P.B. provided the experimental tool for the vocoded sounds. N.G. provided the experimental tool and advice for the streaming segregation task. E.T. helped for the recruitment, the inclusion and the experiments with cochlear implant users. B.T. and A.C designed the experiment, provided statistical analysis and critical revision. All authors discussed the results and implications and commented on the manuscript at all stages.

7. Funding

This work was supported by the ‘‘Fédération pour la Recherche sur le Cerveau (FRC) ”. This work was conducted in the framework of the LabEx CeLyA (‘‘Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) and of the LabEx Cortex (‘‘Construction, Function and 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).

8. References

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*. https://doi.org/10.3389/fnins.2019.01165

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports*, *6*, 18861. https://doi.org/10.1038/srep18861

Albouy, P., Lévêque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., & Caclin, A. (2015). Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia*, *67*, 111–120. https://doi.org/10.1016/j.neuropsychologia.2014.12.006

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: A Journal of Neurology*, *136*(Pt 5), 1639–1661. https://doi.org/10.1093/brain/awt082

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., & Caclin, A. (2015). Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data. *Frontiers in Human Neuroscience*, *9*, 20. https://doi.org/10.3389/fnhum.2015.00020

Ambert-Dahan, E., Giraud, A., Sterkers, O., & Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. *Frontiers in Psychology*, *6*, 181. https://doi.org/10.3389/fpsyg.2015.00181

Amieva, H., & Ouvrard, C. (2020). Does Treating Hearing Loss in Older Adults Improve Cognitive Outcomes? A Review. *Journal of Clinical Medicine*, *9*(3), 805. https://doi.org/10.3390/jcm9030805

Barlow, N., Purdy, S. C., Sharma, M., Giles, E., & Narne, V. (2016). The Effect of Short-Term Auditory Training on Speech in Noise Perception and Cortical Auditory Evoked Potentials in Adults with Cochlear Implants. *Seminars in Hearing*, *37*(1), 84–98. https://doi.org/10.1055/s-0035-1570335

Barone, P., Chambaudie, L., Strelnikov, K., Fraysse, B., Marx, M., Belin, P., & Deguine, O. (2016). Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients. *Cortex*, *83*(Supplement C), 259–270. https://doi.org/10.1016/j.cortex.2016.08.005

Böckmann-Barthel, M., Deike, S., Brechmann, A., Ziese, M., & Verhey, J. L. (2014). Time course of auditory streaming: do CI users differ from normal-hearing listeners? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00775

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Bugannim, Y., Roth, D. A.-E., Zechoval, D., & Kishon-Rabin, L. (2019). Training of Speech Perception in Noise in Pre-Lingual Hearing Impaired Adults With Cochlear Implants Compared With Normal Hearing Adults. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *40*(3), e316–e325. https://doi.org/10.1097/MAO.0000000000002128

Busch, T., Vanpoucke, F., & van Wieringen, A. (2017). Auditory Environment Across the Life Span of Cochlear Implant Users: Insights From Data Logging. *Journal of Speech,*

Language, and Hearing Research: JSLHR, *60*(5), 1362–1377. https://doi.org/10.1044/2016_JSLHR-H-16-0162

Butera, I. M., Stevenson, R. A., Mangus, B. D., Woynaroski, T. G., Gifford, R. H., & Wallace, M. T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. *Scientific Reports*, *8*(1), 11345. https://doi.org/10.1038/s41598-018-29598-x

Caclin, A., Bouchet, P., Djoulah, F., Pirat, E., Pernier, J., & Giard, M.-H. (2011). Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Research*, *1396*, 35–44. https://doi.org/10.1016/j.brainres.2011.04.016

Chari, D. A., Barrett, K. C., Patel, A. D., Colgrove, T. R., Jiradejvong, P., Jacobs, L. Y., & Limb, C. J. (2020). Impact of Auditory-Motor Musical Training on Melodic Pattern Recognition in Cochlear Implant Users. *Otology & Neurotology*, *41*(4), e422. https://doi.org/10.1097/MAO.0000000000002525

Cheng, X., Liu, Y., Shu, Y., Tao, D.-D., Wang, B., Yuan, Y., Galvin, J. J., Fu, Q.-J., & Chen, B. (2018). Music Training Can Improve Music and Speech Perception in Pediatric Mandarin-Speaking Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216518759214. https://doi.org/10.1177/2331216518759214

Choi, J. E., Moon, I. J., Kim, E. Y., Park, H.-S., Kim, B. K., Chung, W.-H., Cho, Y.-S., Brown, C. J., & Hong, S. H. (2017). Sound Localization and Speech Perception in Noise of Pediatric Cochlear Implant Recipients: Bimodal Fitting Versus Bilateral Cochlear Implants. *Ear and Hearing*, *38*(4), 426–440. https://doi.org/10.1097/AUD.0000000000000401

Claes, A. J., Heyning, P. V. de, Gilles, A., Rompaey, V. V., & Mertens, G. (2018). Cognitive outcomes after cochlear implantation in older adults: A systematic review. *Cochlear Implants International*, *0*(0), 1–16. https://doi.org/10.1080/14670100.2018.1484328

Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., & Zatorre, R. J. (2019). The Music-In-Noise Task (MINT): A Tool for Dissecting Complex Auditory Perception. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00199

Cooper, H. R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. *Hearing Research*, *225*(1), 11–24. https://doi.org/10.1016/j.heares.2006.11.010

Cooper, H. R., & Roberts, B. (2009). Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition. *The Journal of the Acoustical Society of America*, *126*(4), 1975. https://doi.org/10.1121/1.3203210

Deroche, M. L. D., Felezeu, M., Paquette, S., Zeitouni, A., & Lehmann, A. (2019). Neurophysiological Differences in Emotional Processing by Cochlear Implant Users, Extending Beyond the Realm of Speech. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000701

Döge, J., Baumann, U., Weissgerber, T., & Rader, T. (2017). Single-sided Deafness: Impact of Cochlear Implantation on Speech Perception in Complex Noise and on Auditory Localization Accuracy. *Otology & Neurotology*, *38*(10). https://doi.org/10.1097/MAO.0000000000001520

Everhardt, M. K., Sarampalis, A., Coler, M., Baskent, D., & Lowie, W. (2020). Meta-Analysis on the Identification of Linguistic and Emotional Prosody in Cochlear Implant Users and Vocoder Simulations. *Ear and Hearing*, *Publish Ahead of Print*. https://doi.org/10.1097/AUD.0000000000000863

Firestone, G. M., McGuire, K., Liang, C., Zhang, N., Blankenship, C. M., Xiang, J., & Zhang, F. (2020). A Preliminary Study of the Effects of Attentive Music Listening on Cochlear Implant Users' Speech Perception, Quality of Life, and Behavioral and Objective

Measures of Frequency Change Detection. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00110

Fiveash, A., Schön, D., Canette, L.-H., Morillon, B., Bedoin, N., & Tillmann, B. (2020). A stimulus-brain coupling analysis of regular and irregular rhythms in adults with dyslexia and controls. *Brain and Cognition*, *140*, 105531. https://doi.org/10.1016/j.bandc.2020.105531

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, *17*(9), 1442–1452. https://doi.org/10.1162/0898929054985446

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific Brain Networks during Explicit and Implicit Decoding of Emotional Prosody. *Cerebral Cortex*, *22*(5), 1107–1117. https://doi.org/10.1093/cercor/bhr184

Fuller, C., Başkent, D., & Free, R. (2019). Early Deafened, Late Implanted Cochlear Implant Users Appreciate Music More Than and Identify Music as Well as Postlingual Users. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.01050

Fuller, C. D., Galvin, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of Two Music Training Approaches on Music and Speech Perception in Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216518765379. https://doi.org/10.1177/2331216518765379

Galvin, 3rd JJ, Fu, Q. J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners., Melodic Contour Identification by Cochlear Implant Listeners. *Ear and Hearing, Ear and Hearing*, *28, 28*(3, 3), 302, 302–319. https://doi.org/10.1097/01.aud.0000261689.35445.20, 10.1097/01.aud.0000261689.35445.20

Galvin, 3rd JJ, Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users., Melodic Contour Identification and Music Perception by Cochlear Implant Users. *Annals of the New York Academy of Sciences, Annals of the New York Academy of Sciences*, *1169, 1169*, 518, 518–533. https://doi.org/10.1111/j.1749-6632.2009.04551.x, 10.1111/j.1749-6632.2009.04551.x

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J.-C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *The Journal of the Acoustical Society of America*, *124*(5), 3076–3087. https://doi.org/10.1121/1.2988289

Giannantonio, S., Polonenko, M. J., Papsin, B. C., Paludetti, G., & Gordon, K. A. (2015). Experience Changes How Emotion in Music Is Judged: Evidence from Children Listening with Bilateral Cochlear Implants, Bimodal Devices, and Normal Hearing. *PloS One*, *10*(8), e0136685. https://doi.org/10.1371/journal.pone.0136685

Glennon, E., Svirsky, M. A., & Froemke, R. C. (2020). Auditory cortical plasticity in cochlear implant users. *Current Opinion in Neurobiology*, *60*, 108–114. https://doi.org/10.1016/j.conb.2019.11.003

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of Music Training for Perception of Emotional Speech Prosody in Deaf Children With Cochlear Implants. *Ear and Hearing*, *38*(4), 455–464. https://doi.org/10.1097/AUD.0000000000000402

Grasso, P. A., Làdavas, E., & Bertini, C. (2016). Compensatory Recovery after Multisensory Stimulation in Hemianopic Patients: Behavioral and Neurophysiological Components. *Frontiers in Systems Neuroscience*, *10*, 45. https://doi.org/10.3389/fnsys.2016.00045

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital

amusia. *Brain and Cognition*, *136*, 103614. https://doi.org/10.1016/j.bandc.2019.103614

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, *111*(3), 1340–1348.

Hirel, C., Nighoghossian, N., Lévêque, Y., Hannoun, S., Fornoni, L., Daligault, S., Bouchet, P., Jung, J., Tillmann, B., & Caclin, A. (2017). Verbal and musical short-term memory: Variety of auditory disorders after stroke. *Brain and Cognition*, *113*(Supplement C), 10–22. https://doi.org/10.1016/j.bandc.2017.01.003

Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *The Journal of the Acoustical Society of America*, *120*(1), 360–374.

Hong, R. S., & Turner, C. W. (2009). Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. *The Journal of the Acoustical Society of America*, *126*(1), 291–299. https://doi.org/10.1121/1.3140592

Hopyan, T., Manno, F. A. M., Papsin, B. C., & Gordon, K. A. (2016). Sad and happy emotion discrimination in music by children with cochlear implants. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *22*(3), 366–380. https://doi.org/10.1080/09297049.2014.992400

Hoth, S., Rösli-Khabas, M., Herisanu, I., Plinkert, P. K., & Praetorius, M. (2016). Cochlear implantation in recipients with single-sided deafness: Audiological performance. *Cochlear Implants International*, *17*(4), 190–199. https://doi.org/10.1080/14670100.2016.1176778

Huang, J., Sheffield, B., Lin, P., & Zeng, F.-G. (2017). Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise. *Scientific Reports*, *7*. https://doi.org/10.1038/s41598-017-02429-1

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

Innes-Brown, H., Marozeau, J., & Blamey, P. (2011). The Effect of Visual Cues on Difficulty Ratings for Segregation of Musical Streams in Listeners with Impaired Hearing. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0029327

Jiam, N. T., & Limb, C. J. (2019). Rhythm processing in cochlear implant-mediated music perception. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.14130

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355.

Kaandorp, M. W., Smits, C., Merkus, P., Festen, J. M., & Goverts, S. T. (2017). Lexical-Access Ability and Cognitive Predictors of Speech Recognition in Noise in Adult Cochlear Implant Users. *Trends in Hearing*, *21*, 2331216517743887. https://doi.org/10.1177/2331216517743887

Kuroda, T., Ono, F., & Kadota, H. (2017). [Dynamic Attending Binds Time and Rhythm Perception]. *Brain and Nerve = Shinkei Kenkyu No Shinpo*, *69*(11), 1195–1202. https://doi.org/10.11477/mf.1416200895

Lê, S., Rennes, A., Josse, J., Rennes, A., Husson, F., & Rennes, A. (2008). FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, 1–18.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge Core - Psychology Research Methods and Statistics. https://doi.org/10.1017/CBO9781139087759

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. *Auditory Cognitive Neuroscience*, 343. https://doi.org/10.3389/fnins.2015.00343

Lerousseau, J. P., Hidalgo, C., & Schön, D. (2020). Musical Training for Auditory Rehabilitation in Hearing Loss. *Journal of Clinical Medicine*, *9*(4), 1058. https://doi.org/10.3390/jcm9041058

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*. https://doi.org/10.1038/srep34911

Lo, C. Y., McMahon, C. M., Looi, V., & Thompson, W. F. (2015). Melodic Contour Training and Its Effect on Speech in Noise, Consonant Discrimination, and Prosody Perception for Cochlear Implant Recipients. *Behavioural Neurology*, *2015*, 352869. https://doi.org/10.1155/2015/352869

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology*, *64*, 109–143. https://doi.org/10.1159/000094648

Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: results from congenital amusia. *Frontiers in Psychology*, *6*, 1340. https://doi.org/10.3389/fpsyg.2015.01340

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-Perception Mismatch in Tone-Deafness. *Current Biology : CB*, *18*(8), R331–R332. https://doi.org/10.1016/j.cub.2008.02.045

Luo, X., Masterson, M. E., & Wu, C.-C. (2014). Contour identification with pitch and loudness cues using cochlear implants. *The Journal of the Acoustical Society of America*, *135*(1), EL8-14. https://doi.org/10.1121/1.4832915

Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). The acoustic and perceptual cues affecting melody segregation for listeners with a cochlear implant. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00790

Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., & Deguine, O. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hearing Research*, *275*(1), 120–129. https://doi.org/10.1016/j.heares.2010.12.010

McAdams, S. (1989). Segregation of concurrent sounds. I: Effects of frequency modulation coherence. *The Journal of the Acoustical Society of America*, *86*(6), 2148–2159.

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification*, *8*(2), 49–82. https://doi.org/10.1177/108471380400800203

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Mosnier, I., Bebear, J.-P., Marx, M., Fraysse, B., Truy, E., Lina-Granade, G., Mondain, M., Sterkers-Artières, F., Bordure, P., Robier, A., Godey, B., Meyer, B., Frachet, B., Poncet-Wallet, C., Bouccara, D., & Sterkers, O. (2015). Improvement of cognitive function after cochlear implantation in elderly patients. *JAMA Otolaryngology-- Head & Neck Surgery*, *141*(5), 442–450. https://doi.org/10.1001/jamaoto.2015.129

Nie, Y., & Nelson, P. B. (2015). Auditory stream segregation using amplitude modulated bandpass noise. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01151

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

Pak, C. L., & Katz, W. F. (2019). Recognition of emotional prosody by Mandarin-speaking adults with cochlear implants. *The Journal of the Acoustical Society of America*, *146*(2), EL165–EL171. https://doi.org/10.1121/1.5122192

Paquette, S., Ahmed, G. D., Goffi-Gomez, M. V., Hoshino, A. C. H., Peretz, I., & Lehmann, A. (2018). Musical and vocal emotion perception for cochlear implants users. *Hearing Research*. https://doi.org/10.1016/j.heares.2018.08.009

Paredes-Gallardo, A., Madsen, S. M. K., Dau, T., & Marozeau, J. (2018a). The Role of Place Cues in Voluntary Stream Segregation for Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216517750262. https://doi.org/10.1177/2331216517750262

Paredes-Gallardo, A., Madsen, S. M. K., Dau, T., & Marozeau, J. (2018b). The Role of Temporal Cues in Voluntary Stream Segregation for Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216518773226. https://doi.org/10.1177/2331216518773226

Passamonti, C., Bertini, C., & Làdavas, E. (2009). Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia*, *47*(2), 546–555. https://doi.org/10.1016/j.neuropsychologia.2008.10.008

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech Intonation Perception Deficits in Musical Tone Deafness (congenital Amusia). *Music Perception: An Interdisciplinary Journal*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

Pattisapu, P., Lindquist, N. R., Appelbaum, E. N., Silva, R. C., Vrabec, J. T., & Sweeney, A. D. (2020). A Systematic Review of Cochlear Implant Outcomes in Prelingually-deafened, Late-implanted Patients. *Otology & Neurotology*, *Publish Ahead of Print*. https://doi.org/10.1097/MAO.0000000000002555

Peng, S.-C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing*, *29*(3), 336–351. https://doi.org/10.1097/AUD.0b013e318168d94d

Petersen, B., Andersen, A. S. F., Haumann, N. T., Højlund, A., Dietz, M. J., Michel, F., Riis, S. K., Brattico, E., & Vuust, P. (2020). The CI MuMuFe – A New MMN Paradigm for Measuring Music Discrimination in Electric Hearing. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00002

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

Pralus, Agathe, Lévêque, Y., Blain, S., Bouchet, P., Tillmann, B., & Caclin, A. (2019). Training pitch perception in cochlear implant users via audiovisual integration: a pilot study in normal-hearing controls. *Implants&Music Congress, Montréal 2019*, *Poster presentation*.

Riley, P. E., Ruhl, D. S., Camacho, M., & Tolisano, A. M. (2018). Music Appreciation after Cochlear Implantation in Adult Patients: A Systematic Review. *Otolaryngology–Head and Neck Surgery*, *158*(6), 1002–1010. https://doi.org/10.1177/0194599818760559

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research*, *1188*, 87–99. https://doi.org/10.1016/j.brainres.2007.10.049

Rouger, J., Lagleyre, S., Démonet, J., Fraysse, B., Deguine, O., & Barone, P. (2012). Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. *Human Brain Mapping*, *33*(8), 1929–1940. https://doi.org/10.1002/hbm.21331

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7295–7300. https://doi.org/10.1073/pnas.0609419104

Saki, N., Nikakhlagh, S., Mirmomeni, G., & Bayat, A. (2019). Perceptual organization of sequential stimuli in cochlear implant listeners: A temporal processing approach. *The International Tinnitus Journal*, *23*(1), 37–41. https://doi.org/10.5935/0946-5448.20190007

Sato, T., Yabushita, T., Sakamoto, S., Katori, Y., & Kawase, T. (2019). In-home auditory training using audiovisual stimuli on a tablet computer: Feasibility and preliminary results. *Auris Nasus Larynx*. https://doi.org/10.1016/j.anl.2019.09.006

Sharma, S. D., Cushing, S. L., Papsin, B. C., & Gordon, K. A. (2020). Hearing and speech benefits of cochlear implantation in children: A review of the literature. *International Journal of Pediatric Otorhinolaryngology*, *133*, 109984. https://doi.org/10.1016/j.ijporl.2020.109984

Sharp, A., Delcenserie, A., & Champoux, F. (2018). Auditory Event-Related Potentials Associated With Music Perception in Cochlear Implant Users. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00538

Shirvani, S., Jafari, Z., Motasaddi Zarandi, M., Jalaie, S., Mohagheghi, H., & Tale, M. R. (2016). Emotional Perception of Music in Children With Bimodal Fitting and Unilateral Cochlear Implant. *The Annals of Otology, Rhinology, and Laryngology*, *125*(6), 470–477. https://doi.org/10.1177/0003489415619943

Shirvani, S., Jafari, Z., Sheibanizadeh, A., Motasaddi Zarandy, M., & Jalaie, S. (2014). Emotional perception of music in children with unilateral cochlear implants. *Iranian Journal of Otorhinolaryngology*, *26*(77), 225–233.

Spangmose, S., Hjortkjær, J., & Marozeau, J. (2019). Perception of Musical Tension in Cochlear Implant Listeners. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00987

Stewart, L. (2011). Characterizing congenital amusia. *Quarterly Journal of Experimental Psychology (2006)*, *64*(4), 625–638. https://doi.org/10.1080/17470218.2011.552730

StreInikov, K., Marx, M., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2015). PET-imaging of brain plasticity after cochlear implantation. *Hearing Research*, *322*, 180–187. https://doi.org/10.1016/j.heares.2014.10.001

StreInikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Démonet, J.-F., Déguine, O., & Barone, P. (2015). Increased audiovisual integration in cochlear-implanted deaf patients: independent components analysis of longitudinal positron emission tomography data. *The European Journal of Neuroscience*, *41*(5), 677–685. https://doi.org/10.1111/ejn.12827

Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology*, *50*(5), 437–444. https://doi.org/10.1111/j.1467-9450.2009.00741.x

Strelnikov, K., Rouger, J., Demonet, J., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain: A Journal of Neurology*, *136*(Pt 12), 3682–3695. https://doi.org/10.1093/brain/awt274

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handbook of Clinical Neurology*, *129*, 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia*, *85*, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640, Part B*, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

Tillmann, B., Poulin-Charronnat, B., Gaudrain, E., Akhoun, I., Delbé, C., Truy, E., & Collet, L. (2019). Implicit Processing of Pitch in Postlingually Deafened Cochlear Implant Users. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01990

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

Timm, L., Vuust, P., Brattico, E., Agrawal, D., Debener, S., Büchner, A., Dengler, R., & Wittfoth, M. (2014). Residual neural processing of musical sound features in adult cochlear implant users. *Frontiers in Human Neuroscience*, *8*, 181. https://doi.org/10.3389/fnhum.2014.00181

Torppa, R., Faulkner, A., Huotilainen, M., Jaervikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children: The role of auditory working memory and musical activities. *INTERNATIONAL JOURNAL OF AUDIOLOGY*, *53*(3), 182–191.

Torppa, R., & Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hearing Research*, *380*, 108–122. https://doi.org/10.1016/j.heares.2019.06.003

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*.

Völter, C., Götze, L., Dazert, S., Falkenstein, M., & Thomas, J. P. (2018). Can cochlear implantation improve neurocognition in the aging population? *Clinical Interventions in Aging*, *13*, 701–712. https://doi.org/10.2147/CIA.S160517

Waddington, E., Jaekel, B. N., Tinnemore, A. R., Gordon-Salant, S., & Goupell, M. J. (2020). Recognition of Accented Speech by Cochlear-Implant Listeners: Benefit of Audiovisual Cues. *Ear and Hearing*, *Publish Ahead of Print*. https://doi.org/10.1097/AUD.0000000000000842

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., Kesteren, E.-J. van, Doorn, J. van, Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1–19. https://doi.org/10.3758/s13423-017-1323-7

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., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory (Hove, England)*, *18*(6), 657–669. https://doi.org/10.1080/09658211.2010.501339

Wright, R., & Uchanski, R. M. (2012). Music Perception and Appraisal: Cochlear Implant Users and Simulated CI Listening. *Journal of the American Academy of Audiology*, *23*(5), 350–379. https://doi.org/10.3766/jaaa.23.5.6

Zhou, Q., Gu, X., & Liu, B. (2019). [The music quality feeling and music perception of adult cochlear implant recipients]. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi =*

Journal of Clinical Otorhinolaryngology, Head, and Neck Surgery, *33*(1), 47–51. https://doi.org/10.13201/j.issn.1001-1781.2019.01.010

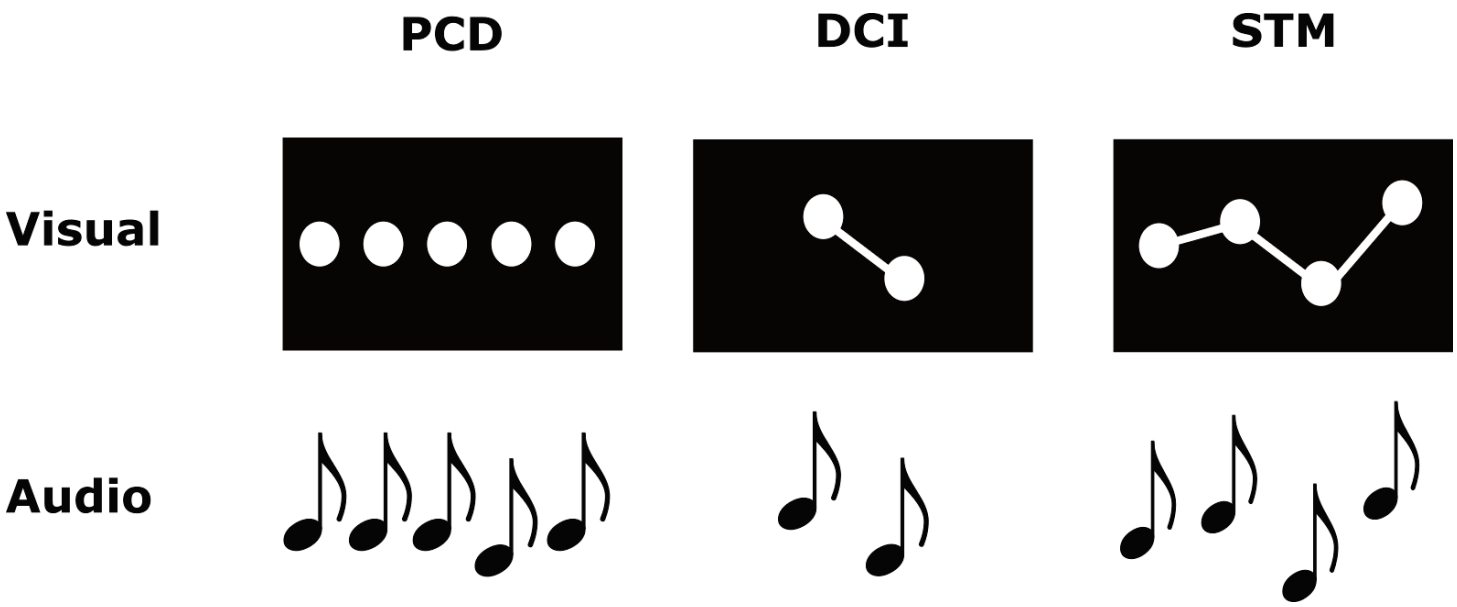

Figure 1 : Audiovisual trials in the Pitch Change Detection (PCD), Direction Change Identification (DCI), and Short-Term Memory (STM) tasks. *The top panels present the visual representations that participants see on the tablet screen in audiovisual trials. Bottom panels present a visual representation of the tones played simultaneously to the visual information. Note that visual stimuli (disks) appears one at a time, simultaneously with a tone, and remain on the screen during the rest of the stimulation (PCD, DCI, STM), as well as during the retention delay before S2 (STM).*

Figure2

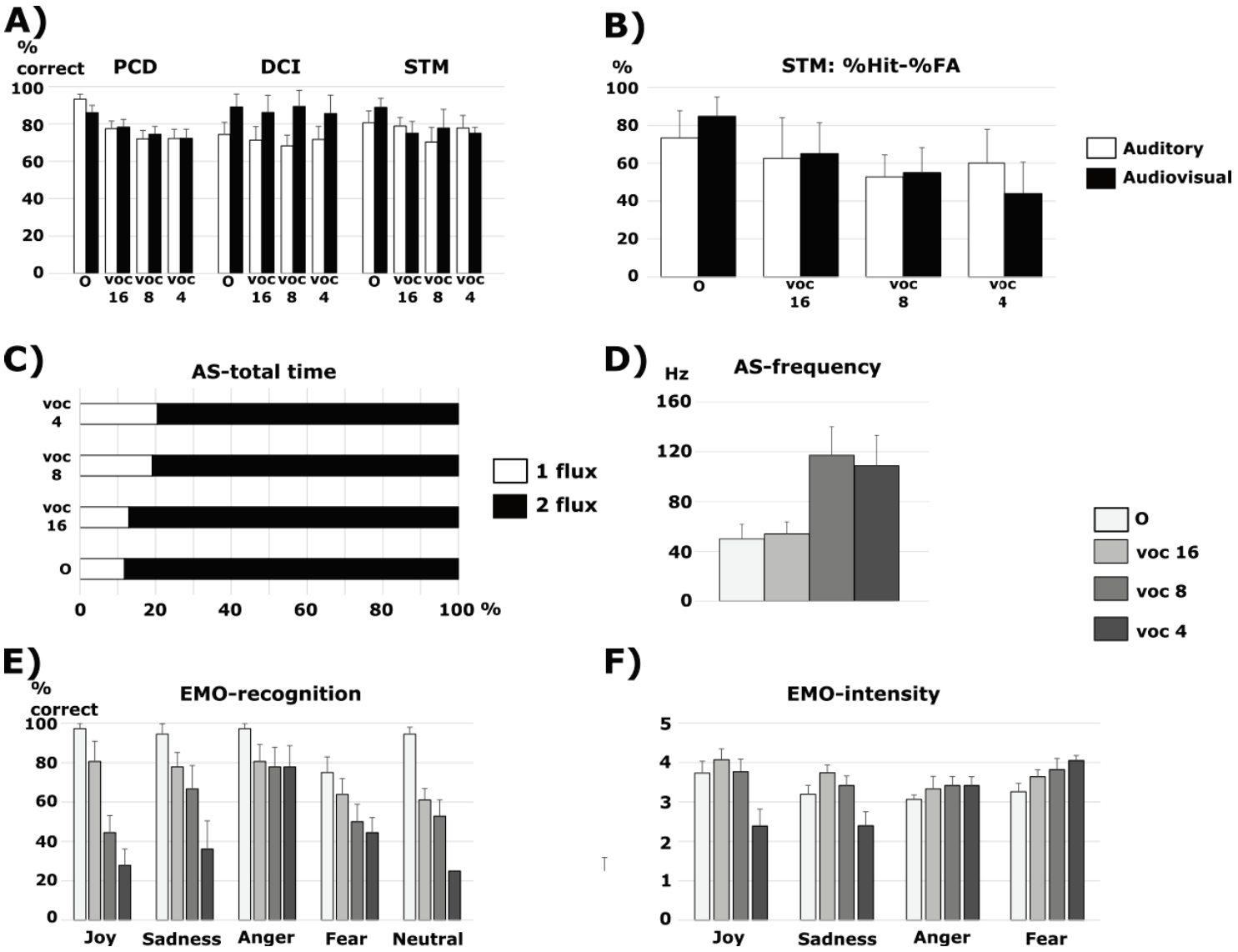

Figure 2: results of the NH participants with original (O) and vocoded sounds (16, 8 or 4 channels) in the five tasks (PCD, DCI, STM, AS and EMO). *For PCD, DCI and STM, percentage of correct responses for auditory (white bars) and audiovisual (black bars) trials are reported (A). For STM, an additional analysis of percentage of Hits minus percentage of False Alarm (FA) was performed (B). For AS, the total time (in percentage) in the percept of 1 (in white) or 2 (in black) flux is reported (C), as well as the mean frequency at change of percept (in Hertz) (D). For EMO, percentage of correct recognition (E) and mean intensity ratings (F) are reported for each emotion. Error bars represent standard error of the mean. PCD: Pitch Change Detection, DCI: Direction Change Identification, STM: Short-Term Memory, AS: Auditory Streaming, EMO: Emotional prosody.*

Figure3

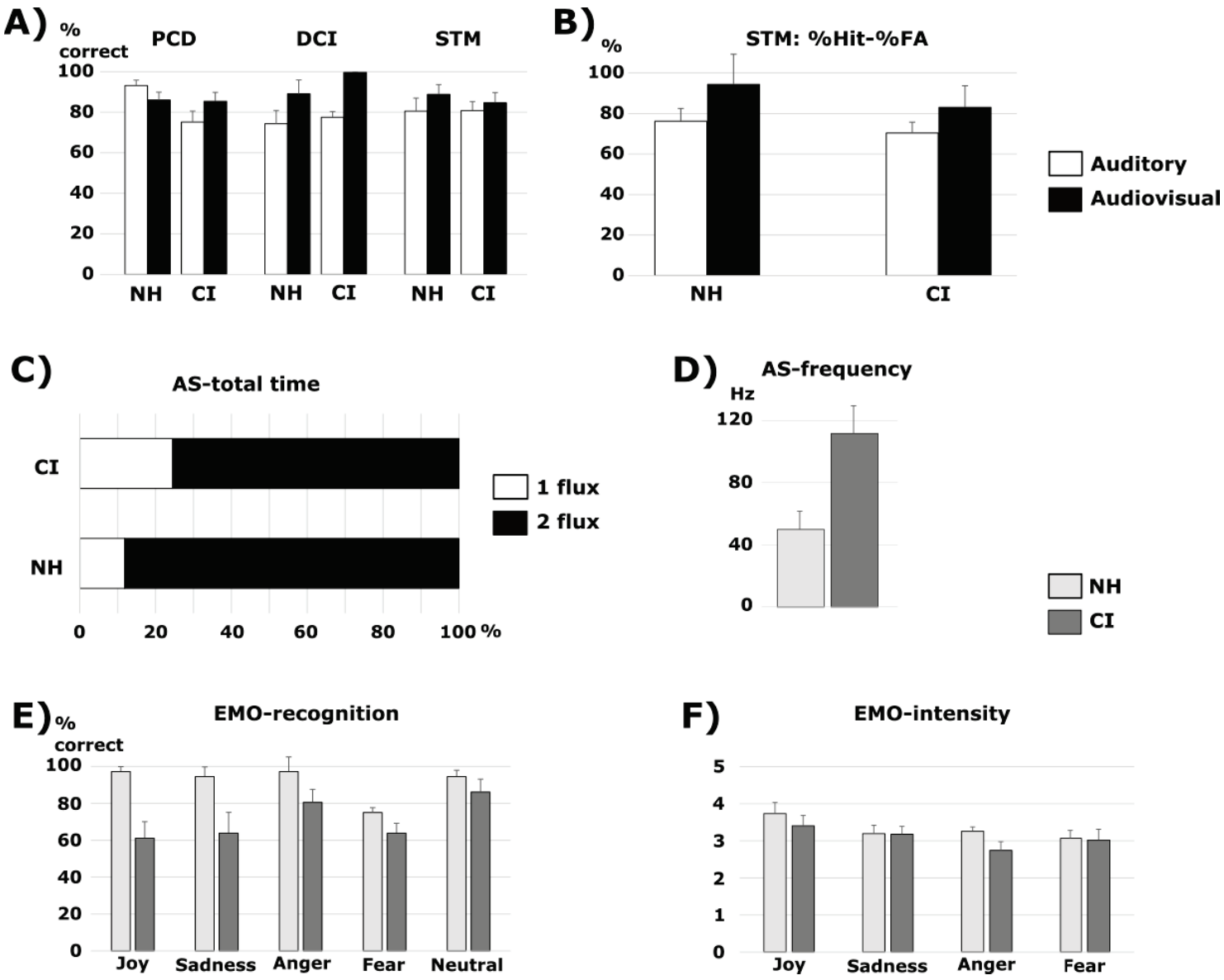

Figure 3: results of the NH participants (with original sounds) and CI users in the five tasks (PCD, DCI, STM, AS, and EMO). *For PCD, DCI and STM, percentage of correct responses for auditory (white bars) and audiovisual (black bars) trials are reported (A). For STM, an additional analysis of percentage of Hits minus percentage of False Alarm (FA) was performed (B). For AS, the total time (in percentage) in the percept of 1 (in white) or 2 (in black) flux is reported (C), as well as the mean frequency at change of percept (in Hertz) (D). For EMO, percentage of correct recognition (E) and mean intensity ratings (F) are reported for each emotion. Error bars represent standard error of the mean. PCD: Pitch Change Detection, DCI: Direction Change Identification, STM: Short-Term Memory, AS: Auditory Streaming, EMO: Emotional prosody.*

Figure4

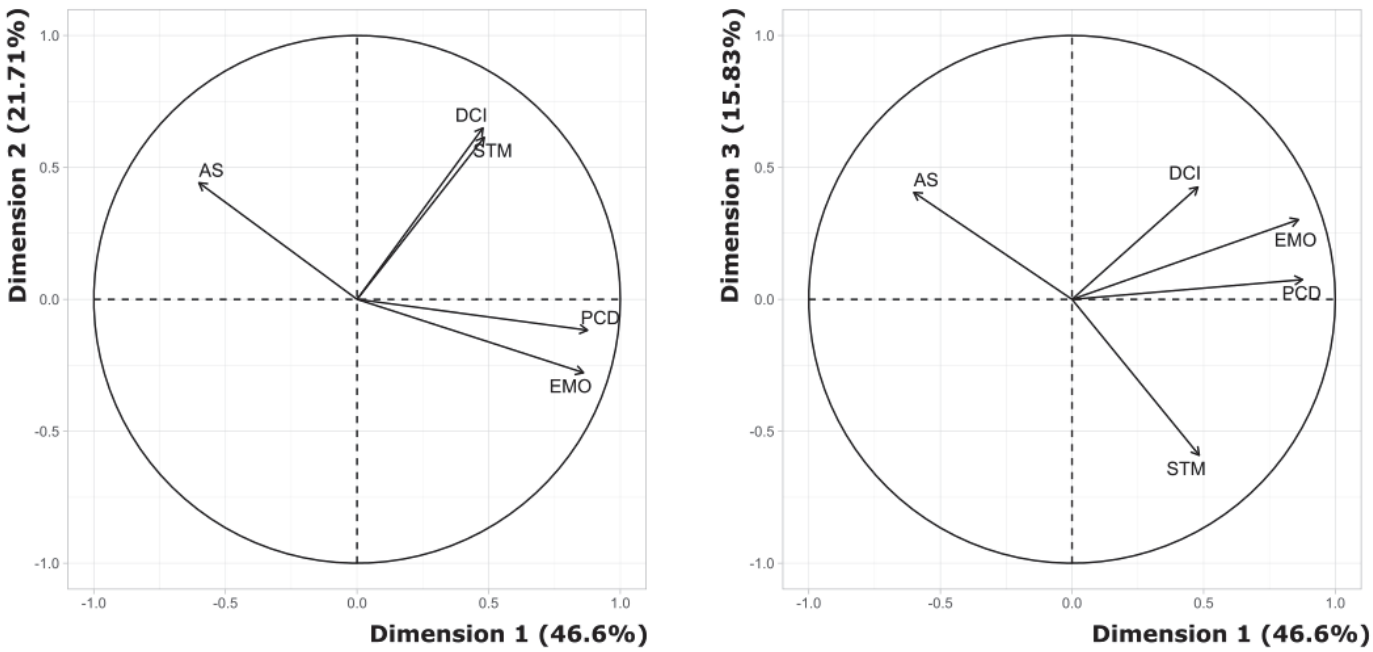

Figure 4: Results of the ACP on the five tasks, across the twenty participants (NH, n = 10; CI users, n = 10). *Three dimensions explained 84% of the variance. PCD, EMO and AS correlated with the first dimension. STM and DCI correlated with the second dimension. STM correlated with the third dimension. PCD: Pitch change Detection - audio trials, DCI: Direction Change Identification - audio trials, STM: Short-Term Memory - audio trials, AS: Auditory Streaming - total time spent in the percept of one flux, EMO: Emotional prosody - recognition.*

Group	*CI* (6 unilateral and 4 bilateral)	*Controls*	*P-value* (group comparison)
Sex	8M 2F	4M 6F	0.07
Age (years)	51 (±14) Min: 24 Max: 73	22.1(±1.7) Min: 20 Max: 25	<0.001
Education (years)	16.1 (±2.8) Min: 10 Max: 20	15.5 (±1.2) Min: 14 Max: 17	0.5
Musical education (years)	1.5 (±4.7) Min: 0 Max: 15	0.6 (±1.6) Min: 0 Max: 5	0.6
Laterality	9D, 1G	9D, 1G	1
Right Ear	8 implants, 2 hearing-aids	NA	
Left Ear	6 implants, 4 hearing-aids	NA	
Unilateral implant (n=6): Duration (years)	2.33 (±1.5) Min: 1 Max: 5	NA	
Bilateral implants (n=4): First implant Duration (years)	6.75 (±6.4) Min: 2 Max: 16	NA	
Bilateral implants (n=4): Second implant Duration (years)	5 (±4.5) Min: 1 Max: 11	NA	

Table 1: demographic data of participants (CI and controls). *The standard deviation is indicated in parentheses. Groups were compared with t.tests (two sided), except for sex where a Chi2 test was used (Qobs= 3.3).*

	Models	*P(M)*	*P(M|data)*	*BF $_M$*	*BF $_{10}$*	*error %*
PCD	Null model (incl. subject)	0.2	1.073e -8	4.290e -8	1.000	
	Sound Type	**0.2**	**0.694**	**9.092**	**6.475e +7**	**0.610**
	Sound Type + Modality	0.2	0.196	0.973	1.824e +7	3.178
	Sound Type + Modality + Sound Type ✻ Modality	0.2	0.110	0.494	1.024e +7	1.274
	Modality	0.2	2.644e -9	1.058e -8	0.247	1.176
DCI	Null model (incl. subject)	0.2	3.164e -7	1.266e -6	1.000	
	Modality	**0.2**	**0.881**	**29.605**	**2.784e +6**	**5.315**
	Sound Type + Modality	0.2	0.099	0.438	311979.244	1.818
	Sound Type + Modality + Sound Type ✻ Modality	0.2	0.020	0.083	64215.941	1.075
	Sound Type	0.2	3.109e -8	1.244e -7	0.098	1.218
STM	Null model (incl. subject)	0.2	0.561	5.105	1.000	
	Sound Type	**0.2**	**0.197**	**0.982**	**0.351**	**0.524**
	Modality	0.2	0.134	0.619	0.239	1.331
	Sound Type + Modality	0.2	0.055	0.235	0.099	10.172
	Sound Type + Modality + Sound Type ✻ Modality	0.2	0.053	0.223	0.094	2.392
AS-total time	Null model (incl. subject)	0.2	2.668e -49	1.067e -48	1.000	
	Sound Type + Percept + Sound Type ✻ Percept	**0.2**	**0.875**	**28.028**	**3.280e +48**	**3.159**
	Percept	0.2	0.116	0.523	4.332e +47	1.046
	Sound Type + Percept	0.2	0.009	0.038	3.484e +46	1.184
	Sound Type	0.2	1.883e -50	7.532e -50	0.071	2.129

AS-frequency	Null model (incl. subject)	0.5	0.073	0.079	1.000	
	Sound Type	**0.5**	**0.927**	**12.720**	**12.720**	**0.307**
EMO-recognition	Null model (incl. subject)	0.2	1.560e -20	6.240e -20	1.000	
	Sound Type + Emotion + Sound Type ✻ Emotion	**0.2**	**0.886**	**31.106**	**5.680e +19**	**0.682**
	Sound Type + Emotion	0.2	0.114	0.514	7.301e +18	0.934
	Sound Type	0.2	6.305e -5	2.522e -4	4.042e +15	0.600
	Emotion	0.2	3.105e -19	1.242e -18	19.904	0.658
EMO-intensity	Null model (incl. subject)	0.2	3.605e -6	1.442e -5	1.000	
	Sound Type + Emotion + Sound Type ✻ Emotion	**0.2**	**0.999**	**4038.321**	**277105.29**	**0.859**
	Sound Type + Emotion	0.2	8.423e -4	0.003	233.632	0.864
	Sound Type	0.2	1.304e -4	5.218e -4	36.182	0.789
	Emotion	0.2	1.320e -5	5.280e -5	3.661	0.550

Table 2: Results of the Bayesian mixed repeated measures ANOVAs on each task (PCD, STM, DCI, AS, EMO), comparing NH participants for 4 Sound types (original, vocoded 4, 8 and 16 channels). *Best model for each task is in bold font. P(M): prior probability assigned to the model, P(M|data): probability of the model knowing the data, BF_M: Bayesian Factor of the model, BF_{10}: Bayesian Factor of the model compared to the null model.*

	Models	*P(M)*	*P(M|data)*	*BF $_M$*	*BF $_{10}$*	*error %*
PCD	Null model (incl. subject)	0.2	0.036	0.149	1.000	
	Modality + Group + Modality ∗ Group	**0.2**	**0.898**	**35.147**	**24.947**	**9.508**
	Group	0.2	0.040	0.165	1.099	1.790
	Modality + Group	0.2	0.014	0.058	0.395	1.472
	Modality	0.2	0.012	0.050	0.345	1.591
DCI	Null model (incl. subject)	0.2	1.111e -4	4.444e -4	1.000	
	Modality	**0.2**	**0.464**	**3.465**	**4178.231**	**1.488**
	Modality + Group	0.2	0.301	1.721	2708.463	1.336
	Modality + Group + Modality ∗ Group	0.2	0.235	1.227	2113.544	1.824
	Group	0.2	5.756e -5	2.303e -4	0.518	0.660
STM	Null model (incl. subject)	0.2	0.325	1.928	1.000	
	Modality	**0.2**	**0.307**	**1.774**	**0.945**	**1.368**
	Group	0.2	0.151	0.714	0.466	0.658
	Modality + Group	0.2	0.147	0.690	0.452	2.488
	Modality + Group + Modality ∗ Group	0.2	0.069	0.296	0.212	2.282
AS-total time	Null model (incl. subject)	0.2	1.367e -17	5.469e -17	1.000	
	Percept + Group + Percept ∗ Group	**0.2**	**0.878**	**28.803**	**6.422e +16**	**1.531**
	Percept	0.2	0.090	0.394	6.558e +15	0.953
	Percept + Group	0.2	0.032	0.133	2.360e +15	1.629
	Group	0.2	4.768e -18	1.907e -17	0.349	1.044

AS-frequency	Null model	0.5	0.163	0.195	1.000	
	Group	**0.5**	**0.837**	**5.125**	**5.125**	**7.765e -4**
EMO-recognition	Null model (incl. subject)	0.2	2.095e -4	8.380e -4	1.000	
	Emotion + Group + Emotion ✻ Group	**0.2**	**0.860**	**24.653**	**4107.699**	**2.294**
	Emotion + Group	0.2	0.123	0.563	588.756	0.879
	Emotion	0.2	0.014	0.058	67.967	0.347
	Group	0.2	0.002	0.007	8.747	2.277
EMO-intensity	Null model (incl. subject)	0.2	0.295	1.673	1.000	
	Emotion	**0.2**	**0.315**	**1.838**	**1.067**	**0.787**
	Emotion + Group	0.2	0.174	0.843	0.590	0.956
	Group	0.2	0.161	0.769	0.547	0.868
	Emotion + Group + Emotion ✻ Group	0.2	0.055	0.233	0.187	1.238

Table 3 Results of the Bayesian mixed repeated measures ANOVAs on each task (PCD, STM, DCI, AS, EMO), comparing NH participants and CI users (Group). *Best model for each task is in bold font. P(M): prior probability assigned to the model, P(M|data): probability of the model knowing the data, BF_M: Bayesian Factor of the model, BF_{10}: Bayesian Factor of the model compared to the null model.*

	Answered / Expected	*Joy*	*Sadness*	*Anger*	*Fear*	*Neutral*
Original						
	Joy	**97.2**	0	0	0	2.8
	Sadness	0	**94.4**	2.8	2.8	0
	Anger	0	0	**97.2**	2.8	0
	Fear	2.8	16.7	2.8	**75**	2.8
	Neutral	0	2.8	2.8	0	**94.4**
Vocoded 16						
	Joy	**82.5**	7.5	2.5	5	2.5
	Sadness	7.5	**80**	0	7.5	5
	Anger	0	0	**82.5**	15	2.5
	Fear	12.5	5	15	**67.5**	0
	Neutral	0	22.5	5	7.5	**65**
Vocoded 8						
	Joy	**47.5**	7.5	12.5	25	7.5
	Sadness	0	**65**	2.5	15	17.5
	Anger	0	0	**80**	17.5	2.5
	Fear	2.5	10	27.5	**52.5**	7.5
	Neutral	5	22.5	2.5	10	**60**
Vocoded 4						
	Joy	**27.8**	5.6	33.3	19.4	13.9
	Sadness	2.8	**36.1**	22.2	13.9	25
	Anger	2.8	0	**77.8**	11.1	8.3
	Fear	5.6	5.6	33.3	**44.4**	11.1

Neutral	0	38.9	19.4	16.7	**25**

Table 4: Percent of answer types for each intended emotion for EMO task, averaged over the 10 NH participants for each Sound Types (Original, vocoded 16, 8 or 4 channels). *Correct answers are on the diagonal.*

	Answered / Expected	*Joy*	*Sadness*	*Anger*	*Fear*	*Neutral*
NH participants						
	Joy	**97.2**	0	0	0	2.8
	Sadness	0	**94.4**	2.8	2.8	0
	Anger	0	0	**97.2**	2.8	0
	Fear	2.8	16.7	2.8	**75**	2.8
	Neutral	0	2.8	2.8	0	**94.4**
CI users						
	Joy	**57.5**	20	5	15	2.5
	Sadness	7.5	**62.5**	0	5	25
	Anger	2.5	0	**77.5**	12.5	7.5
	Fear	0	7.5	20	**62.5**	10
	Neutral	0	15	0	0	**85**

Table 5: Percent of answer types for each intended emotion for EMO task, averaged over the 10 participants for each group (NH participants and CI users). *Correct answers are on the diagonal.*

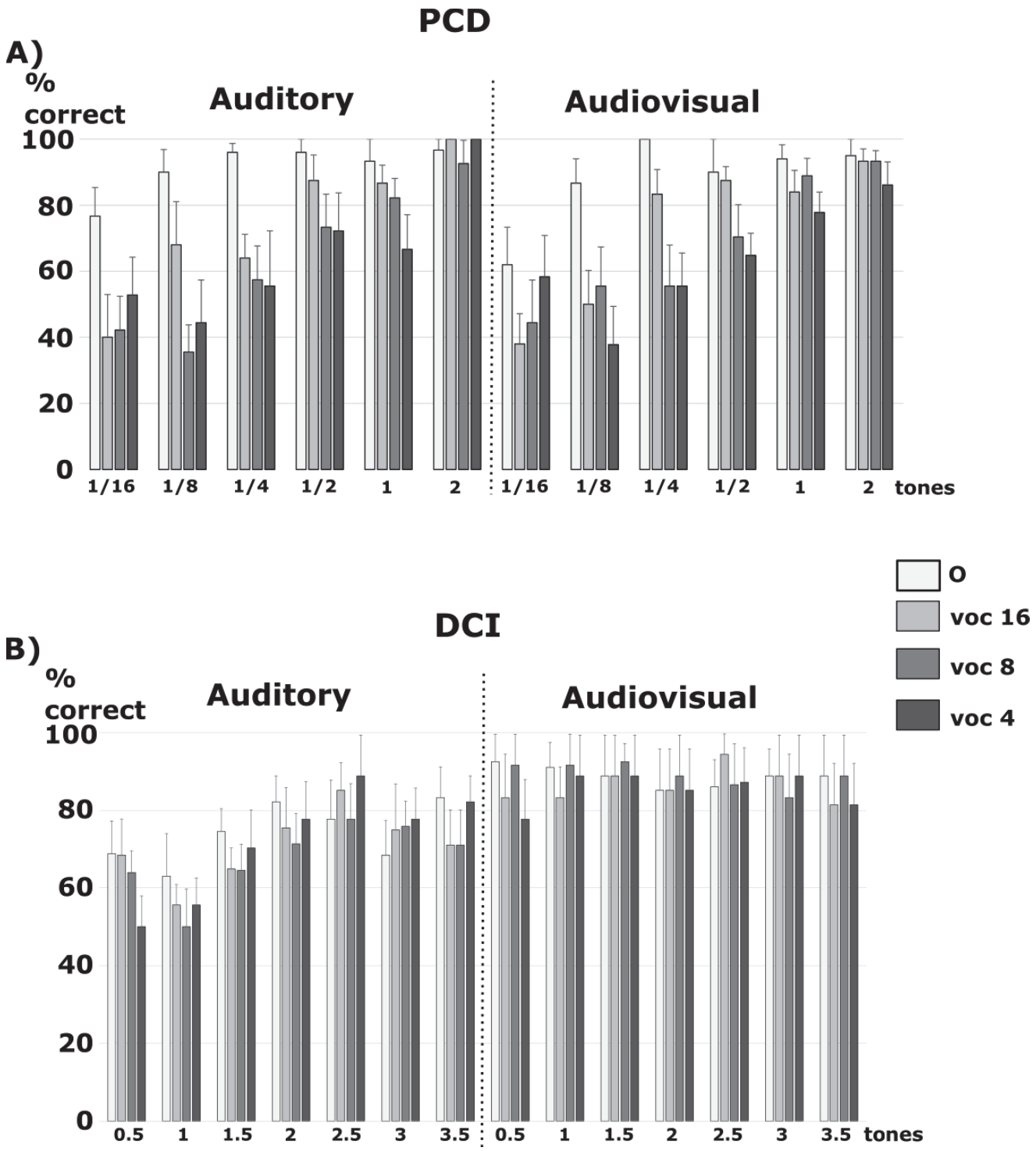

Figure S1: results of the NH participants for original (O) and vocoded sounds (16, 8 and 4 channels) in the PCD (different trials) and DCI, according to size changes. *Percentage of correct responses for auditory and audiovisual trials are reported for PCD (A) and DCI (B). For PCD, six size changes were present for the different trials: 1/16, 1/8, ¼, ½, 1 or 2 tones. For DCI, seven size changes were presented: 0.5, 1, 1.5, 2; 2.5, 3, or 3.5 tones. Error bars represent standard error of the mean. PCD: Pitch change Detection, DCI: Direction Change Identification.*

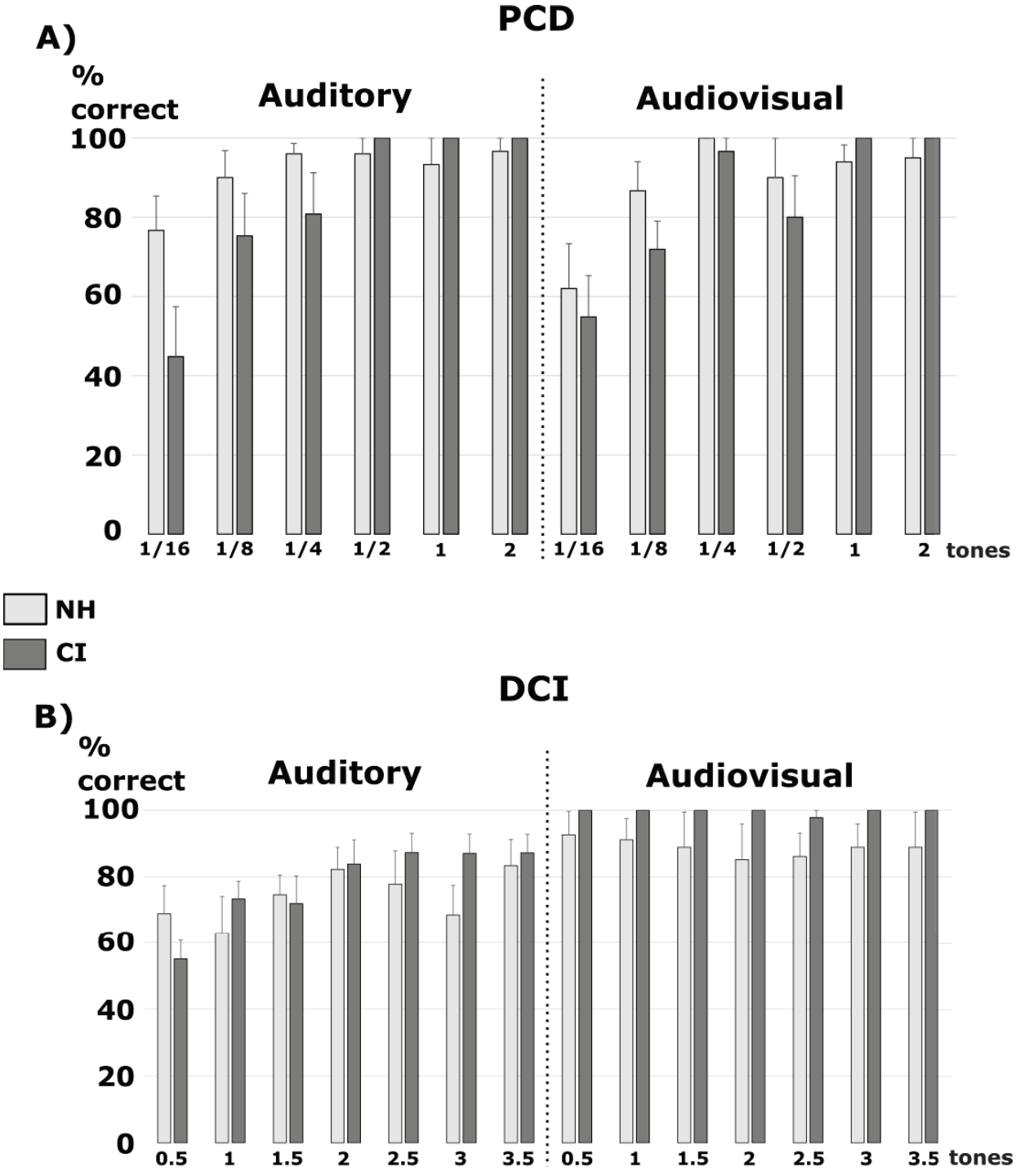

Figure S2: results of the NH participants and CI users in the PCD and DCI, according to size changes. *Percentage of correct responses for auditory and audiovisual trials are reported for PCD (A) and DCI (B). For PCD, six size changes were present for the different trials: 1/16, 1/8, ¼, ½, 1 or 2 tones. For DCI, seven size changes were presented: 0.5, 1, 1.5, 2; 2.5, 3, or 3.5 tones. Error bars represent standard error of the mean. PCD: Pitch change Detection, DCI: Direction Change Identification.*

B. Study 5: Building a training program to enhance non-verbal auditory abilities (in progress)

1. Introduction

We designed a new rehabilitation program using the three audiovisual tasks previously described in Study 4: a pitch-change detection task, a pitch direction change identification task, and a short-term memory for pitch task. For these three tasks, informative or non-informative visual cues were available. As the training program was adaptive, each participant could enhance its performance at its own level. As a control for this training procedure, we also implemented a control training using visuo-spatial stimuli, previously developed by Nathalie Bedoin (Bedoin & Medina, 2013). Based on a previous study investigating the feasibility of training for CI users (Sato et al., 2019), we aimed to keep the training as convenient as possible for the participants. To do so, each participant was given a touch tablet to do the training at home. This training is composed of two training sessions per week of 15 minutes each, for fifteen weeks. Indeed, as this training is designed for further use by CI users as a rehabilitation perspective, we wanted to keep it feasible in an everyday routine, hence no too overwhelming for participants.

In NH participants, we also aim to test the direct effect of this audiovisual training on brain plasticity. To do so, we recorded magnetoencephalography measures with two tasks. The first task was a short-term memory task with tones, similar to the one in the training program. The second task was a passive oddball paradigm with pitch deviants. Source reconstruction should allow to detect brain plasticity changes occurring after the training session, especially in the fronto-temporal network.

2. *Material and Methods*

a. Protocol design

Participants are included for a total of 45 weeks, with two types of training sessions for 15 weeks and 15 weeks without training (see **figure 1**). In total, four testing sessions are planned in the laboratory with 15 weeks between each testing session, before and after the training sessions or the control period without training (see **figure 1**). At each testing session (from 1 to 4), participants are tested with the testing battery (Study 4) to assess their auditory and attentional abilities, before and after each type of training session (audiovisual or visuospatial), and after the control period. Moreover, control participants undergo MEG recordings, as due to the implants, CI users could not be recorded with MEG[1]. At Testing Session 1, each participant is given a touch tablet (Ipad) to perform each training. Between the testing session 1 and 2, and testing session 2 and 3, participants performed one training (audiovisual or visuospatial) then the other one. A randomization between participants is done for the training order realization. Trainings are performed at home, 2 times per week, participant are contacted regularly to make sure the training is done correctly. For control participants, an anatomical MRI is recorded on testing session 1 (3T Siemens Magnetom Prisma, CERMEP), and magnetoencephalographic recordings are performed on testing sessions 1, 2 and 3.

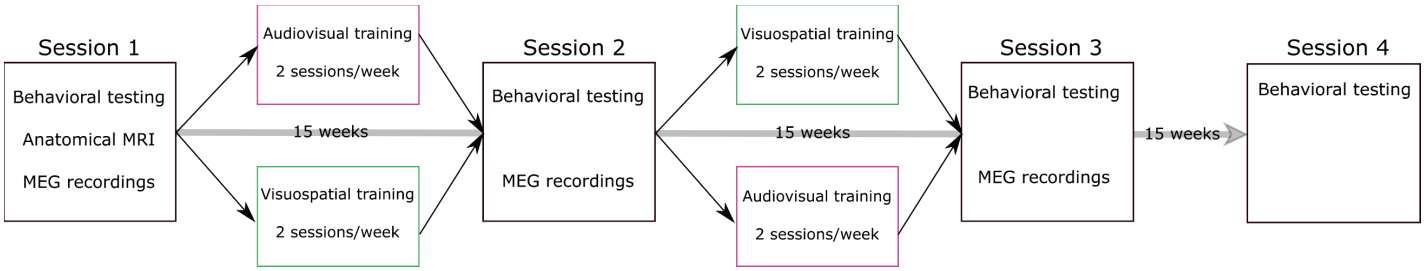

Figure 1: Training procedure for all participants. In control participants, MEG and MRI are recorded, in CI users, only behavioral measures are collected. Each participant is included in the entire procedure (45 weeks) and participates to the two trainings consecutively. Trainings order is counter-balanced among participants.

[1] We plan to recruit congenital amusics as well to perform the trainings.

b. *Audiovisual training: Material and procedure*

The audio-visual training is composed of three subtests: Pitch Change Detection (PCD), pitch Direction Change Identification (DCI) and pitch Short-Term Memory (STM). All tests are implemented to run on an iPad touch tablet, allowing participants to answer by touching large buttons presented on the screen.

For PCD, DCI, and STM tasks, the same stimuli are used, with a roving of frequency across trials. They are synthetic harmonic tones (twelve harmonics), equalized in Root Mean Square (RMS) amplitude, each lasting 500 ms and presented with a within-sequence Inter-Stimulus-Interval (ISI) of 100 ms. On half of the trials in PCD, DCI and STM tasks, visual cues are presented. They are composed of white disks with a diameter of 2 cm (100 px) on a black screen. At home, the participant is asked to seat in front of the tablet with two loudspeakers (Logitech Z200) at 70 centimeters of distance from the participant's head, with 40 centimeters between each speaker. Participants can adjust the volume to a comfortable listening level.

For each training session, each participant performs the three subtests in a random order, with each test presenting the stimuli in pseudo-random order, with no more than two repetitions of the same type of stimulus (same pitch) in a row. Before each subtest, participants receives a visual explanation of the corresponding task with a written support. One training session lasts about 15 minutes.

Over the training period, the difficulty level of each subtest is adapted to the participant's level (between 1: low difficulty level, and 20: high difficulty level). The difficulty level is based on the pitch change size, the bigger the pitch change, the easier the test is (see **table** for details). Each participant starts at a level of 5 for the three subtests. If the participant has a score higher than 80% of correct responses over two consecutive sessions, the level is raised by one level (see **table 1**). If the participant has a score lower than 50% of correct responses over two consecutive sessions, the level is lowered by one level.

i. *Pitch Change Detection (PCD) test*

In one trial, participants are presented with a sequence of five isochronous tones, all identical (standard tone) except the fourth tone that could differ in frequency (adapted from Hyde & Peretz, 2004, and Albouy et al., 2015). Standard frequencies are 165, 196, 220, 262, 330, 392, 494, 659, 784, 880, 1047,

or 1319Hz. Deviant frequencies are between 131 and 1662Hz, with changes relative to the standard tone being between 1/16, 1/8, ¼, ½, 1, 1.25, 1.5, 1.75 or 2 tones, either up or down compared to the standard. For each session, 64 sequences are constructed with five notes. There are 16 identical trials (four trials per each standard) and 48 different trials (twelve trials per each standard, that is one trial per deviant size, up and down). Non-informative visual cues are presented on half of the trials in addition to the tones (as in Albouy et al. (2015)). Five circles appear sequentially from left to right, with the onset of appearance of a circle being synchronized to the onset of each tone. They are always positioned at the center of the vertical axis on the screen, and hence are not informative as far as the pitch of the tone is concerned but give information about the temporal moment of the onset of the sound. Participants have to determine if the fourth tone of the sequence is the same as or different from the other tones. After the end of the sequence, participants have unlimited time to give their answer by tapping on either the "Same" or the "Different" button. After having given their answer, the next trial is played automatically after an average delay of 1000ms (700-1300ms).

ii. Pitch Direction Change Identification (DCI) test

For each trial, participants are presented with two tones at two different frequencies. Fundamental frequencies of the tones are comprised between 123 and 2093 Hz. Steps between the two tones can be of 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 or 5 tones. For each session, 56 pairs are constructed with two different tones, 28 "up" sequences, that is the second tone being higher in pitch than the first one, 28 "down" sequences, that is the second tone being lower in pitch than the first one. Informative visual cues are presented on half of the trials in addition to the tones. Two circles connected by a white bar are appearing consecutively and simultaneously with the onset of each tone. Circles' vertical positions are centered on the average position of the frequency of the two tones (to be at the center of the screen) and are calculated according the frequency of the corresponding tone, the higher the frequency, the higher the circle on the screen. In contrast to PCD, visual cues are thus fully informative for pitch height, but as they are only present in half of the trials, participants are asked to base their judgements on their auditory perception. These visual cues aim to reinforce the association between visual height and pitch (Pralus, Lévêque et al., 2019). Participants have to determine if the second tone is higher in pitch (Up) or lower (Down) than

the first tone. After the end of the second tone, participants have unlimited time to give their answer by tapping on either the "Up" or the "Down" button. After having given their answer, the next trial ss played automatically after an average delay of 1000ms (700-1300ms).

iii. Pitch Short-Term Memory (STM) test

Difficulty level	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PCD Pitch change (tones)	1.75 and 2	1.5 and 1.75	1.5 and 1.75	1.25 and 1.5	1.25 and 1.5	1 and 1.25	1 and 1.25	1 and 1.25	1/2 and 1	1/2 and 1	1/2 and 1	1/4 and 1/2	1/4 and 1/2	1/4 and 1/2	1/8 and 1/4	1/8 and 1/4	1/8 and 1/4	1/16 and 1/8	1/16 and 1/8	1/16 and 1/8
DCI Pitch change (tones)	4, 4.5 and 5	3.5 and 4	3.5 and 4	3, 3.5 and 4	3, 3.5 and 4	2.5, 3 and 3.5	2.5, 3 and 3.5	2.5, 3 and 3.5	2 and 2.5	2 and 2.5	2 and 2.5	1.5, 2 and 2.5	1.5, 2 and 2.5	1.5, 2 and 2.5	1 and 1.5	1 and 1.5	0.5 and 1	0.5 and 1	0.5 and 1	0.5 and 1
STM Pitch change (tones)	4.5 and 5	4	3.5	3 and 3.5	3	2.5 and 3	2.5 and 3	2.5	2 and 2.5	2 and 2.5	2	1.5 and 2	1.5 and 2	1.5	1 and 1.5	1 and 1.5	1	0.5 and 1	0.5 and 1	0.5

***Table 1**: Change sizes (tones) for each audiovisual task (PCD, DCI and STM) according to difficulty level.*

Participants are presented with two melodies of four tones (S1 and S2), with S2 being either identical or different from S1 (adapted from Tillmann et al., 2009; and Hirel et al., 2017). Fundamental frequencies of the tones are comprised between 131 and 523 Hz (corresponding to notes between C3 and C5). For each session, 32 melodies are constructed with four tones, each melody lasts 2300 ms. In total, there are 16 identical and 16 different trials. For different trials, changes of one tone can occur on the second or third tone. Changes can be of 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 or 5 tones, all entailing a change of contour. The delay between the two melodies of a trial is 1000ms. Informative visual cues are presented on half of the trials in addition to the tones, they are presented during S1 and the delay between S1 and S2. Circles are connected by white bars, appearing consecutively and simultaneously with the onset of each tone of the first sequence. Circles' vertical positions are centered on average over the four tones, each

vertical circle position is calculated according to the frequency of the corresponding tone, the higher the frequency, the higher the circle would be on the screen. Participants have to determine if the second melody is the same or different from the first melody. After the end of the second melody, participants have unlimited time to give their answer by tapping on "Same" or "Different" button. After giving their answer, the next trial is played automatically after an average delay of 1000ms (700-1300ms).

c. *Visuospatial training: material and procedure*

The training is composed of five subtests: arrow-global, arrow-alternance, triplet-global, triplet-local, triplet-complex (see **figure 2**). All tests are implemented to run on an iPad touch tablet, allowing participants to answer by touching large buttons presented on the screen. These five tasks were previously used in children with language disorder for remediation of visuo-spatial disorders (see Switchipido from Bedoin, 2017; Bedoin & Medina, 2013 for complete descriptions of the tasks).

For each training session, each participant performs the five subtests in a random order, with each test presenting the stimuli in pseudo-random order, with no more than two repetitions of the same type of stimulus (same answer) in a row. Before each subtest, participants receive a visual explanation of the corresponding task with a written support. One training session lasts about 15 minutes.

i. *Arrow-global*

The displayed stimulus represents a small arrow included in a big arrow, both vertically. If one arrow is white, the other is blue and vice versa. Each arrow can be oriented downwards or upwards to the screen, and both arrows can have the same or different orientation. The participants have to indicate the orientation of the big arrow (up or down) and ignore the orientation of the small arrow. After the appearance of the stimulus on the screen, participants have unlimited time to give their answer by tapping on "Up" or "Down" button, below the stimulus. After giving their answer, the next trial is displayed automatically after an average delay of 1000ms (700-1300ms).

ii. *Arrow-alternance*

The stimulus represents a small arrow included in a big arrow, both vertically and both included in a big circle. If one arrow is white, the other is blue and vice versa, the circle have the same color as the small

arrow. Each arrow can be oriented downwards or upwards to the screen, both arrows can have the same or different orientation. The participants have to indicate the orientation of the white arrow (up or down) and ignore the orientation of the blue arrow. After the appearance of the stimulus on the screen, participants have unlimited time to give their answer by tapping on "Up" or "Down" button, below the stimulus. After giving their answer, the next trial is displayed automatically after an average delay of 1000ms (700-1300ms).

iii. Triplet tasks

One cue symbol is presented on upper middle half of the screen. At the same time, two hierarchized target symbols are presented in the right and left bottom section of the screen. Hierarchized targets are global symbols made of small local symbols (different from the global one). Global and local symbols can represent a heart, a moon, a cross, a cup, or a star. After the appearance of the three stimuli on the screen, participants have unlimited time to give their answer by tapping on one of the two targets. After giving their response, the next trial is displayed automatically after an average delay of 1000ms (700-1300ms).

Global

Participants have to indicate which of the two targets contain the cue by focusing on the global target symbols and ignoring the small local symbols composing these targets.

Local

Participants have to indicate which of the two targets contain the cue by focusing on the global target symbols and the small symbols composing these targets, the cue can be in either of the two targets, in either of the two levels of focalization.

Triplet-complex

The cue is a hierarchized global symbol made of small local symbols (different from the global one). Participants have to indicate which of the two targets contain one of the two symbols composing the cue by focusing on the global target symbols and the small local symbols composing these targets, the cue can be in either of the two targets, in either of the two levels of focalization.

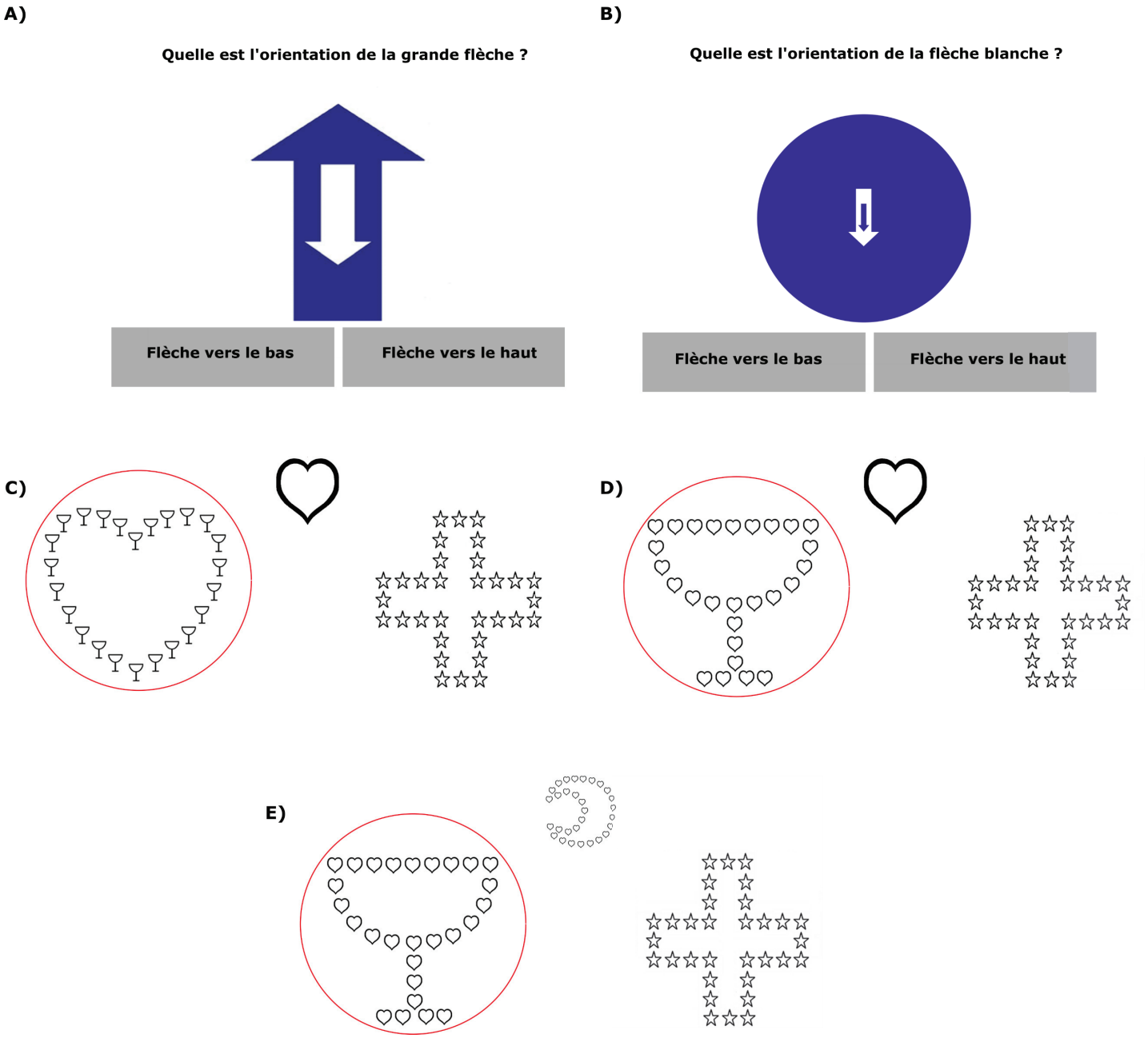

Figure 2: *Visual representations of each subtests of the visuospatial training.* ***A)*** *Arrow-global test, participant has to provide the direction of the big arrow.* ***B)*** *Arrow-alternance test, participant has to give the orientation of the white arrow.* ***C)*** *Triplet global test, participant has to find the cue on the top in the global shape of the two bottom targets.* ***D)*** *Triplet local test, participant has to find the cue on the top in the global or local shapes of the two bottom targets.* ***E)*** *Triplet complex test, participant has to find the cue (global or small symbols) on the top in one of the two bottom targets (global or local shape). Red circles are indicating the correct answer (here for illustration purpose only).*

d. Testing battery

To test the effect of each of the two trainings in the performance of the participants, we use a testing behavioral battery to assess their abilities in auditory and visuospatial cognition. This battery comprises various subtests:

- The five auditory subtests are described in Study 4 (see STM, PCD, DCI, EMO, and ASA for details), using only auditory stimuli.
- A test assessing hearing in noise (referred to as "audimots" test) (see Moulin & Garcia, 2012; Moulin & Richard, 2015 for full details). For each trial, participants hear a word in noise and have to select the correct answer among four propositions on the screen. Propositions can be similar to the target word phonetically (difficult trials) or very different phonetically (easy trials). Both speech noise and cocktail party noise are used. For control participants, a speech noise ratio (SNR) of -6 dB is used. For CI users, a SNR of 6 and -3 dB re used. The intensity of the signal (words) is fixed at 85 dB.
- The five subtests are the one used in the visuospatial training (see section 3 above for details).
- Testing visuo-attentional abilities with SIGL (see Bedoin, 2017; Kéïta et al., 2014 for details). Hierarchized letters are used as cues, they are global letters made of small letters (different from the global one). In the local subtest of SIGL, these letters are always made of small M or E. In the global subtests of SIGL, these cues are always globally either M or E. Participants have to tell as rapidly as possible if they detected a M or E by tapping on the correct answer either on right or left of the screen. After giving their answer, the next trial is displayed automatically after an average delay of 1000ms (700-1300ms). This task is closed to the visuospatial tasks of the training, and could reflect possible generalization of the training on visuo-attentional abilities of the participants.
- Written texts testing for speed and efficiency of reading (developed by Nathalie Bedoin). The DeltaText test consists in reading a text as fast and as correct as possible, we record reading errors and reading speed. Four different texts are used: one for each testing session. The "Lazy Whale" test consists in copying a text hung on the wall as correctly as possible in 3 minutes

maximum. We record errors of copy and the number of back and forth between the hang text and the copying sheet. This test records reading abilities and memory for words in participants.

- Hearing quality questionnaires recording for hearing difficulties in participants' everyday life (see Alhanbali et al., 2017; Moulin et al., 2019; Moulin & Richard, 2016 for details).

The full testing battery lasted approximately one hour and half. All tests were implemented on touch tablet, except for the questionnaires and the written texts. For each testing session, the subtests were administered in the same order: auditory tests (PCD, DCI, STM, ASA, EMO, audimots), the visuo-attentional tests (Arrow-global, Arrow-alternance, Triplet-Global, Triplet-Local, Triplet-Complexe, SIGL-Local, SIGL-Global), DeltaText, "Lazy Whale" text, Hearing quality questionnaires.

e. Magnetoencephalography recordings

i. Stimuli and procedure

An active Short-Term Memory (STM) test and a passive oddball paradigm are used during MEG recording.

The STM test is implemented similarly as the one used in the testing battery and the audiovisual training. Same sounds are used in the MEG as in the audiovisual training and the testing battery. Only changes of 3 semitones are used. Fundamental frequencies of the tones are comprised between 131 and 523 Hz (corresponding to notes between C3 and C5). Five blocks of 36 trials are presented in the same order for every participant in a given testing session. 180 trials are presented in total (90 different and 90 identical). Participants give their answer by clicking on a left or right button of a mouse (left for "identical", right for "different"). After the end of the S2, they have 3 seconds to give their answer otherwise a series of bip is played to indicate that no response was recorded. After giving their response, the next trial is displayed automatically after an average delay of 2000ms (1700-2300ms). The entire testing lasts about 30min.

For the oddball paradigm, we use similar synthetic harmonic tones than in the testing battery and the audiovisual training, with a duration of 100ms. Two Inter-Stimulus-Interval (ISI) are used (Fakche et al., 2018): two blocks with 400ms and three blocks with 1400ms. For each of the two blocks of 400ms

ISI, there are 840 stimuli (140 deviants). For each of the three blocks of 1400ms ISI, are were 280 stimuli (47 deviants). The standard tone has a fundamental frequency of 262Hz. For one block of 400ms ISI, deviant tone has a fundamental frequency of 266Hz, for the other block, deviant tone has a fundamental frequency of 294hz.For the 1400ms ISI blocks, deviant tone has a fundamental frequency of 294Hz. The entire testing lasts about 35min.

Presentation software (Neurobehavioral systems) is used to present the stimuli and record button presses. The participants start the session with the STM task, breaks are proposed between each block. A practice block with six trials is performed at the beginning to ensure the participant understand the task. No feedback is given during the task. During the oddball paradigm, participants watch a silent movie with subtitles and are told that they do not have not to pay attention to the audio stimuli.

ii. Recordings

A 275-channel whole-head MEG system (CTF-275 by VSM Medtech Inc.) is used for the recordings. We use a continuous sampling rate at 600 Hz, a 0.016-150 bandwidth filter and first-order spatial gradient noise cancellation. We record electrocardiogram as well as horizontal and vertical electrooculograms with bipolar montages. Coils are fixated at the nasion and the preauricular points (fiducial points) to record head position continuously (sampling rate of 150Hz). At the beginning of each block, it is checked that head displacement do not exceed 0.5cm from the starting position. Participants are settled up in a sound-attenuated and magnetically shielded recording room. They are seated in an upright position. They listen to the sounds through air-conducting tubes with foam ear tips. To adjust the level of the sounds individually, participants undergo a sound detection threshold task. We use a tone of 500ms at a fundamental frequency of 262 Hz. The threshold is determined for each ear with an adaptive procedure and we adjust the level at 50dB above the measured threshold.

3. *Preliminary results and interim discussion*

Three control participants started the audiovisual training just before the confinement imposed in France due to the COVID-19 pandemic, we were able to record their behavioral performance at session 2 after 30 training sessions over 15 weeks just before submitting this manuscript. We present here some preliminary results obtained at some tests of the (auditory) testing battery (Study 4), to validate our protocol for further testing with CI users and control participants (**Figure 2**). No MEG recordings could be done in the second testing session (MEG facilities were closed).

***Figure 2**: Results of five subtests of the testing battery, before and after the audiovisual training in three control participants. Each color corresponds to one control participant. For PCD, DCI, STM, and EMO, correct responses are indicated in percentage (left scale). For AS, mean frequency for the change of percept is indicated in hertz (right scale), the lower the frequency is, the better the streaming is.*

For PCD, STM and EMO subtests, no clear change of responses was observed for the three participants before and after the training. Indeed, as their results were already quite high (almost 100% correct responses) before the training, we might have here a ceiling effect for the tasks. However, for DCI and AS subtests we saw a tendency of improved performances after the training. For DCI, we observed higher scores (percentages of correct responses) after the training, revealing a tendency of enhancement in the three participants. Similarly, for AS, the decrease of mean frequency for the change of percept suggest a better streaming capacity in these three subjects.

Overall, these preliminary results suggest that audiovisual training could enhance the auditory perception abilities of control participants, an effect that would be mostly observable for tasks with finer measurements and not at-ceiling performance. However, we expect interesting results with the MEG recordings, in particular for fronto-temporal regions plasticity. Due to the actual context, the second MEG session could not be recorded. We hope to record new control participants as soon as possible to conclude on the efficiency of the audiovisual training program.

General Discussion

VII. Discussion and Perspectives

The research presented in the present thesis focused on non-verbal auditory cognition and potential deficits thereof. This present work was divided in two axes regarding the perception of non-verbal auditory cues. The first axis focused on characterizing emotion perception in two conditions associated with central non-verbal auditory deficits: congenital amusia and after brain lesion. Based on the results of this axis and previous research (Fuller et al., 2019; Stewart et al., 2006; Tillmann et al., 2015), the second axis of this thesis aimed to find a new diagnostic tool and new rehabilitation strategies to measure and enhance non-verbal auditory cognition in several types of deficits. Indeed, numerous researches have aimed for diagnostic tools and rehabilitation tools for verbal cognition, but only very few training strategies exist for the non-verbal cognition domain. Overall, the present research allows for 1) better understanding of specific emotional perception deficits, complementing previous findings reported in this domain (**Table 1**), and 2) draws perspectives for new tools to enhance non-verbal auditory cognition and communication skills of individuals.

Deficit / ***Stimuli***	***Music***	***Emotional prosody***
Congenital Amusia	Previously tested (Lévêque et al., 2018)	Article 1 (Pralus et al., 2019)
Brain lesion	Article 3 (Pralus et al., 2020) Previous case report (Hirel et al., 2014)	Case report (Bourgeois–Vionnet et al., 2020)
CI users	To be tested	Article 4 (Pralus et al., in revision)

Table 1*: Summary of studies using the two-tasks paradigm (recognition and intensity ratings). In this thesis, I report three studies using this paradigm with emotional prosody and musical emotion stimuli in congenital amusia, brain-damaged patients and CI users. I also contributed to a case report using this paradigm with emotional prosody stimuli in a case of brain lesion.*

A. Emotion perception in non-verbal auditory deficits

1. Impaired perception of emotions in music and speech

Studies 1, 2, and 3 of this work reported several deficits regarding emotion perception in music and speech. In Studies 1 and 3, we used the same paradigm testing explicit and implicit perception of emotions. We tested this paradigm in two populations in which we expected impairments: congenital amusics (Study 1) and brain-damaged patients (Study 3). This paradigm was designed to assess potential deficits of both explicit and implicit processing of emotion (Bourgeois–Vionnet et al., 2020; Hirel et al., 2014; Lévêque et al., 2018). The explicit test was a simple forced-choice task to assess emotion recognition by choosing among potential categories. The ratings of the intensity of emotions perceived reflected more implicit processing of emotion with no verbalization needed. These ratings were previously used in an identical paradigm with musical emotions (Hirel et al., 2014; Lévêque et al., 2018). These two studies suggested that this two-tasks paradigm was efficient to detect implicit strategies of emotion processing in participants. Interestingly, in both studies, we observed different results with the explicit and implicit tasks. In congenital amusics (Study 1), even though the recognition task revealed a deficit of emotional prosody perception, the intensity ratings task showed no impairment. In brain-lesioned patients (study 3), LBD patients had a deficit of musical emotion recognition with no intensity ratings deficit, as observed in congenital amusics (Lévêque et al., 2018; Pralus et al., 2019). On the contrary, the RBD patients had preserved emotion recognition whereas they showed altered intensity ratings.

This paradigm allows the detection of preserved implicit perception of emotions despite a deficit of explicit perception. Both in congenital amusics and in LBD patients, it suggests that even if the recognition of emotions is difficult and impaired, some preserved implicit knowledge of emotion could be present. As previously shown in other neuropsychological deficits, implicit knowledge could be spared even if the explicit processing seems disrupted. For example, in participants with aphasia, it was shown that implicit learning was possible with a serial reaction time task, even though their sentence comprehension was disrupted (Schuchard et al., 2017; Schuchard & Thompson, 2014). Hence, this dissociation between implicit and explicit learning opens new avenues for training strategies in these

populations. Indeed, by using trainings with more implicit strategies, we could enhance emotion perception in these populations and overcome the explicit deficit (Tillmann, Lalitte, et al., 2016). Moreover, it gives insights about brain mechanisms involved in emotion perception in these populations, revealing a potential dissociation between networks involved in the feeling of emotions per se, and the networks involved in the retrieval of this information and the conscious access to it. As suggested by Cleeremans and colleagues, this dissociation observed between explicit and implicit mechanisms could reflect a difference in the accessibility of information, related to a "difference in degree, rather than in kind" (Cleeremans & Jiménez, 2002). Thus, research on brain networks involved in consciousness gives further insights of the different states of consciousness and their cerebral correlates (Boly & Seth, 2012). This research would suggest that changes in the posterior and the fronto-temporal cortices could be involved in auditory consciousness (Brancucci et al., 2016; Koch et al., 2016). Hence, we could further investigate the relationships between these brain networks and the emotion perception using our two-task paradigm.

Furthermore, RBD patients were not impaired for the recognition of musical emotions whereas their pattern of intensity ratings was very different from the one of controls or LBD patients. Indeed, RBD patients had low intensity ratings for sadness and fear. First, it confirms a previously observed tendency to under-estimate negative emotions in RBD patients, in keeping with the observation of euphoric mood (Sackeim et al., 1982). Most interestingly, this paradigm has allowed the detection of a specific deficit in RBD patients that could have not been found if we only used explicit recognition tasks. Indeed, as previously suggested by studies with facial and prosodic material in healthy participants and unilateral brain-damaged patients (Abbott et al., 2013; Borod et al., 2002b; Demaree et al., 2005), intensity ratings could reflect the actual experience of emotions. Thus, RBD patients could be more affected in this cognition and not so much with the recognition of emotions.

In Studies 1 and 3, the MBEA and PDT tests revealed that some of the patients had acquired amusia and that MBEA scores correlated with emotion recognition. As MBEA scores reflect the overall quality of the perception of musical stimuli, this finding suggests that recognition of emotions and musical perception abilities are linked, with some participants having a global deficit in evaluating musical

stimuli. In the MBEA evaluation, the participants' memory is needed as most of the subtests required to compare two melodies, suggesting that this test requires more general cognitive abilities to be performed (Särkämö et al., 2009, 2010). Interestingly, in our two studies the poor scores obtained at the MBEA, by both congenital amusics and some of the brain-damaged patients, could reflect some memory issues, in particular regarding pitch. Indeed, three of the LBD patients that had a MMSE score below the cut-off, also had a low MBEA score. Moreover, it was previously suggested that congenial amusics had specific memory deficit for pitch (Albouy, Mattout, et al., 2013; Tillmann et al., 2009; Tillmann, Lévêque, et al., 2016; Williamson & Stewart, 2010). These results suggest that cognitive and perceptual musical abilities are important for explicitly recognizing emotions in music and speech. However, other parameters than musical abilities could be involved for emotion recognition as for example some patients can have acquired amusia without a deficit of emotion recognition (Hirel et al., 2014). Furthermore, congenital amusics can also demonstrate preserved sensitivity to emotional music (Gosselin et al., 2015). We could suggest that implicit knowledge of tonal structure could help these participants for emotion recognition even though explicit knowledge is disrupted (Albouy, Schulze, et al., 2013; Nguyen et al., 2009; Barbara Tillmann, Lalitte, et al., 2016). The variety of profiles observed are in keeping with the hypothesis of (at least partly) separate processes for music perception and emotion (Peretz et al., 1998; Satoh et al., 2011; Stewart et al., 2006). As suggested by imaging studies in healthy participants, emotional response and pleasure felt when listening to music would require the activation of the limbic structures such as the ventral striatum, the amygdala or the orbitofrontal cortex but not the activation of more superior temporal structures, that would be activated for music analysis (Blood et al., 1999; Brown et al., 2004).

Studies 1 and 3 revealed the diversity of patterns of non-verbal auditory cognition deficits, especially regarding emotion perception. It suggested partially distinct brain mechanisms for implicit and explicit emotion processing. As this two-tasks paradigm was powerful to detect specific emotion perception deficits in central disorders (Hirel et al., 2014; Lévêque et al., 2018; Pralus et al., 2019; Pralus et al., 2020), we suggested that it could be useful to detect deficits in peripheral disorders and used it in Study 4 to detect emotional prosody deficit in CI users. It showed that CI users were impaired in explicitly

recognizing joy and sadness in emotional sentences, but their intensity ratings were intact. Thus, intensity ratings, in combination with explicit recognition measures, could allow building a sensitive test to detect possible emotion perception abnormalities in clinical settings, even if a patient is unaware of this deficit (Stewart et al., 2006; Tillmann, Lalitte, et al., 2016).

2. Brain mechanisms of emotion perception

In the first axis of the thesis, we used two different strategies to study brain mechanisms during emotion perception. In congenital amusics, we used electroencephalography to assess automatic mechanisms in the brain during passive listening of emotional prosody (Study 2). In patients, the study of brain lesions localization sheds light on different brain networks involved in musical emotion perception in the two hemispheres (Study 3).

Study 2 revealed that automatic brain processing of neutral and emotional sounds was disrupted in congenital amusics despite some preserved recognition of emotions. In relation with our behavioural study with the same emotional stimuli, we suggested that congenital amusics could have a deficit of awareness for emotional stimuli. However, this deficit could be partly compensated at higher levels of processing as these individuals are still able to perform the recognition task, especially for the anger deviant. However, the categorization task used was rather easy and we could suggest that this compensation would not be sufficient with a more complex explicit task. In Study 3, RBD were still able to recognize emotional stimuli despite their alteration of the emotional feeling of these same stimuli (as reflected by intensity ratings). Overall, these studies 2 and 3 suggest at least partially separate automatic and attentional brain processes for emotion perception. Moreover, it suggests that even if automatic brain processing is disrupted, compensating mechanisms could overcome this deficit for a correct recognition of emotions, at least in optimal conditions. However, the observed brain impairments might lead to additional difficulties to process emotions in some situations. For instance, in degraded conditions such as noisy environment, emotions could be more difficult to perceive.

Study 3 focused on the investigation of the effect of brain-lesion side on the processing of emotions. Interestingly, we found different patterns of alterations according to the side of lesion. We revealed that left-brain damage disrupted emotion recognition whereas right-brain damage affected more strongly the

ratings of intensity of these emotions. This finding suggests a distinction on the cerebral level between cognitive intentional processes of emotion recognition, and the emotional experience of music. This pattern confirmed what was already suggested in studies with healthy participants and unilateral brain-damaged patients with facial and vocal stimuli (Abbott et al., 2013; Borod et al., 2002b; Demaree et al., 2005). These studies suggested that the right hemisphere hypothesis would better apply to emotion recognition, i.e., with a cognitive or intentional process, whereas the valence hypothesis would better apply to automatic processing of the emotion, closer to the emotional experience of participants (Abbott et al., 2013; Borod et al., 2002b; Demaree et al., 2005). These two modes of emotional processing would rely on different anatomical substrates, with emotion recognition associated with hemispheric asymmetries in posterior and temporal regions, whereas emotion experience would be associated with hemispheric asymmetries in more frontal regions (Abbott et al., 2013; Borod, 1992). Here, with the results of study 3 and previous knowledge about brain alterations in congenital amusia (Albouy et al., 2019; Albouy, Mattout, et al., 2013, 2015; Hyde et al., 2011; Leveque et al., 2016; Loui et al., 2009; Tillmann et al., 2015), we gave further insights on these theories with musical material. We could suggest that the implicit processing of emotions could rely more on limbic system and the primary and secondary auditory cortices (Frühholz et al., 2016). Indeed, congenital amusics seem to not have severe alterations of this network as well as the brain-damaged patients. However, the explicit perception of emotions could rely more on frontal regions such as the inferior-frontal cortex and the medial frontal cortex, that seem to be impaired in congenital amusics and brain-damaged patients. Indeed, these networks were demonstrated to be involved in emotional categorization and decoding the relevance of an emotion in a social context (Frühholz & Grandjean, 2012; Hoekert et al., 2008, 2010; Rauschecker & Scott, 2009). Here, studies 1, 2, and 3 give some cues to define the brain networks involved in emotion processing.

In our EEG study with congenital amusics (Study 2), we did not find any clear side effect, however it was previously shown that right fronto-temporal pathways are disrupted in congenital amusia (Albouy et al., 2019; Albouy, Mattout, et al., 2013, 2015; Hyde et al., 2011; Leveque et al., 2016; Loui et al., 2009; Tillmann et al., 2015). Moreover, studies with brain-damaged patients revealed that severe

acquired amusia was more associated to right-hemisphere lesions whereas patients with left-hemisphere lesion had more transient amusia and recovered from it more often (Sihvonen, Ripollés, Särkämö, et al., 2017; Sihvonen, Särkämö, Ripollés, et al., 2017). However, gray matter volume analysis in relation with MBEA performance are still debated as one study demonstrated its relation to the left hemisphere (Mandell et al., 2007) whereas a studies with BD patients demonstrated that amusia could be caused by lesion in both hemisphere (Särkämö et al., 2009; Schuppert et al., 2000; Tillmann et al., 2017), and MBEA score was more related to the right hemisphere (Sihvonen, Ripollés, Särkämö, et al., 2017). Altogether, we could suggest that the two hemispheres are differentially involved in the processing of music, however further investigation is needed to conclude on their exact roles.

B. Training strategy for non-verbal auditory cognition using multisensory integration

1. Assessing non-verbal auditory cognition deficits

As we have seen in the first axis on this thesis, several known deficits could entail troubles regarding auditory cognition that do not relate to the verbal domain. Indeed, in brain-damaged patients for instance, the focus of studies and clinical assessment regarding emotion perception has been mainly on verbal and facial material (Borod et al., 2002a; Charbonneau et al., 2003; Cheung et al., 2006; Harciarek et al., 2006; Yuvaraj et al., 2013), with on rare assessment for music emotion perception (Dellacherie et al., 2011; Gosselin et al., 2011; Jafari et al., 2017; Khalfa et al., 2007). Similarly, congenital amusia has long been described as a music-perception specific deficit (Tillmann et al., 2015) with only recent studies interested in assessing possible deficit regarding music emotion perception (Gosselin et al., 2015; Lévêque et al., 2018). Moreover, research studies have investigated verbal deficit in congenital amusics (Jiang et al., 2010, 2012; Liu et al., 2015, 2017; Nan et al., 2016; Nguyen et al., 2009; Tillmann, Burnham, et al., 2011; Tillmann, Rusconi, et al., 2011), and very few studies were interested in emotional prosody perception (Lima et al., 2016; Lolli et al., 2015; Thompson et al., 2012). In the first axis, we demonstrated that non-verbal cognition is essential to understand and perceive fully our environment and several central and peripheral disorders could suffer from a non-assessed deficit in non-verbal auditory cognition. Moreover, patients with auditory deficits can complain about poor

cognition in the non-verbal auditory domain, despite correct verbal cognition. For example, CI users are able to understand speech in good hearing condition, but they still experience difficulties to perceive speech in noise (Bugannim et al., 2019; Choi et al., 2017; Hong & Turner, 2006) or to enjoy music (Fuller et al., 2019; Riley et al., 2018; Zhou et al., 2019).

As no full assessment battery of non-verbal auditory cognition has yet been designed, we started the second axis of this thesis by designing such a tool. We used tests that have been previously used in studies with healthy participants (Grimault et al., 2002), congenital amusics (Albouy, Lévêque, et al., 2015; Hyde & Peretz, 2004; Tillmann et al., 2009) and brain-damaged patients (Hirel et al., 2017) and adapted them to design a testing battery covering several aspects of non-verbal auditory cognition, with emphasis on pitch perception and memory. This battery assesses rapidly (30 minutes) abilities of participants in pitch change detection, pitch direction change identification, memory for melodies, auditory scene analysis and emotional prosody perception. Interestingly, we demonstrated in Study 4 that this battery can be used easily with both healthy participants and patients, here CI users. We showed that it assessed different aspects of non-verbal auditory cognition as participants did not have the same pattern of results across the different tasks (PCA analysis).

Moreover, we used audiovisual stimuli in three tasks of this battery. This allowed showing that CI users were able to benefit from visual cues to enhance their auditory perception abilities. Indeed, as this was previously suggested in studies using verbal and facial material (Barone et al., 2016; Butera et al., 2018; Rouger et al., 2008), CI users could benefit from visual cues to perceive auditory stimuli, even more than NH participants (Rouger et al., 2007; Strelnikov et al., 2015). Here, we demonstrated with musical material (pitch tones) that this multisensory integration is powerful in CI users, for different types of tasks. We suggest that this audiovisual facilitatory interaction could be used in CI users as a training strategy to enhance their non-verbal auditory perception abilities.

2. Rehabilitation of non-verbal auditory perception and cognition

Based on previous results in CI users (Innes-Brown et al., 2011; Rouger et al., 2007, 2008; Strelnikov et al., 2015), and the results obtained here in Study 4, it seems that audiovisual training could be a good strategy to enhance auditory perception in CI users. However, only a very few number of research has

been done to evaluate the potential benefit of audiovisual training on CI users (Bernstein et al., 2014; Sato et al., 2019; Vandali et al., 2015). These first studies demonstrated that audiovisual training had a positive effect in CI users on speech in noise perception (Bernstein et al., 2014; Sato et al., 2019) and frequency discrimination (Vandali et al., 2015). In Study 5, we designed a new training strategy using audiovisual integration to enhance auditory abilities in CI users. To control for the long-term effect of training, we used a cross-over randomization strategy with a control training using visuo-spatial stimuli. Moreover, as we expect changes in the fronto-temporal networks with our newly developed audiovisual training, we used MEG recordings in the control participants. As this study was delayed due to the health situation in France, we have only collected very preliminary behavioural results in three control participants that did the audiovisual training.

As previous trainings in CI users suggested that the effect of training could be limited (Cheng et al., 2018), we decided to look at the long-term effect of our audiovisual training by recording behavioural results at the assessment battery 15 weeks after the end of the second training. Moreover, we designed the duration of the training (in weeks and number of training sessions per week) according to previous recordings of CI users' feelings and demands (Cullington et al., 2018; Fu & Galvin, 2007). Indeed, the training should be long enough for long-term effect but as it was designed to be usable in CI users' everyday life, only two training sessions a week should be done, to not interfere with their daily routine.

As our training is implemented on touch tablet and easy to realize, we could suggest that it would it possible to adapt it for children CI users. Indeed, no much training strategies has been yet developed in children with CIs, and it could be interesting to investigate this domain as cortical plasticity is strong in children (see Gfeller, 2016 for a review). Indeed, it was shown that CI children that have learnt a music instrument have enhance abilities in non-verbal auditory cognition such as ASA and auditory memory (Canette et al., 2016; Rochette et al., 2014). Some studies have looked at the effect of musical training on auditory abilities in CI children (Abdi et al., 2001; Chen et al., 2010; Cheng et al., 2018; Fu et al., 2015; Good et al., 2017; Rocca, 2012; Rochette & Bigand, 2009; Roman et al., 2016; R. Torppa et al., 2014; Yucel et al., 2009). Overall, they showed some improvements in musical skills, frequency discrimination abilities, speech perception and emotional prosody perception. However, there is still

very little knowledge about the effect of musical training in the long term in children CI users (Rochette & Bigand, 2009). Interestingly, children with CIs seem to have the willing to involve in musical activities and enjoy listening to music (Gfeller et al., 1999), even more than adults, suggesting that a training for non-verbal auditory cognition very early in life for these children would be appreciated.

VIII. Conclusion

This thesis investigated non-verbal auditory cognition in central and peripheral deficits. The first part of this work focused on emotion perception in central disorders, it demonstrated a dissociation between implicit and explicit processing for the perception of emotions in speech and music, in congenital amusia and brain-damaged patients. Moreover, by using behavioural and electroencephalographic measures, this first section gives further insights on brain networks involved in emotion perception and its correlation with musical perception abilities. Based on these studies and previous knowledge, this thesis has allowed for the design of a new training strategy for non-verbal auditory cognition using multisensory integration, opening perspectives for rehabilitation. Overall, it provides new testing and training tools running on touch tablet available for diverse populations (normal-hearing participants, patients with central auditory deficits, patients with peripheral hearing loss) to assess non-verbal auditory deficits and to enhance non-verbal cognition over time.

Further experiments need to be run using this new training tool to assess its efficiency directly and in the long-term. The MEG recordings planned during this training will give new insights about the cerebral correlates of plasticity of non-verbal auditory cognition. Moreover, even if we focused our research on CI users, this tool is easy to adapt to other patients' populations and was designed to be feasible at-home. Hence, it could be developed to be used in several populations like congenital amusics in which no existing training strategies is yet developed. Indeed, previous work on congenital amusia as well as this present work have shown that it is a deficit still not perfectly understood, especially regarding brain mechanisms underlying congenital amusia. It would be interesting to record brain networks during our pitch perception and memory tasks to further characterize their functioning as well as investigate whether an audiovisual training could give rehabilitation on the long-term for this population.

References

Abbott, J. D., Cumming, G., Fidler, F., & Lindell, A. K. (2013). The perception of positive and negative facial expressions in unilateral brain-damaged patients: A meta-analysis. *Laterality*, *18*(4), 437–459. https://doi.org/10.1080/1357650X.2012.703206

Abdi, S., Khalessi, M. H., Khorsandi, M., & Gholami, B. (2001). Introducing music as a means of habilitation for children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *59*(2), 105–113. https://doi.org/10.1016/S0165-5876(01)00460-8

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*. https://doi.org/10.3389/fnins.2019.01165

Albouy, P., Lévêque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., & Caclin, A. (2015). Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia*, *67*, 111–120. https://doi.org/10.1016/j.neuropsychologia.2014.12.006

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: A Journal of Neurology*, *136*(Pt 5), 1639–1661. https://doi.org/10.1093/brain/awt082

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., & Caclin, A. (2015). Altered retrieval of melodic information in congenital amusia: Insights from dynamic causal modeling of MEG data. *Frontiers in Human Neuroscience*, *9*, 20. https://doi.org/10.3389/fnhum.2015.00020

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

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults. *Ear and Hearing*, *38*(1), e39–e48. https://doi.org/10.1097/AUD.0000000000000361

Ambert-Dahan, E., Giraud, A., Sterkers, O., & Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. *Frontiers in Psychology*, *6*, 181. https://doi.org/10.3389/fpsyg.2015.00181

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Atiani, S., David, S. V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S. A., & Fritz, J. B. (2014). Emergent Selectivity for Task-Relevant Stimuli in Higher-Order Auditory Cortex. *Neuron*, *82*(2), 486–499. https://doi.org/10.1016/j.neuron.2014.02.029

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusiaA group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. https://doi.org/10.1093/brain/awf028

Balkwill, L.-L., & Thompson, W. F. (1999). A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64. JSTOR. https://doi.org/10.2307/40285811

Ball, T., Rahm, B., Eickhoff, S. B., Schulze-Bonhage, A., Speck, O., & Mutschler, I. (2007). Response Properties of Human Amygdala Subregions: Evidence Based on Functional MRI Combined with Probabilistic Anatomical Maps. *PLoS ONE*, *2*(3). https://doi.org/10.1371/journal.pone.0000307

Barlow, N., Purdy, S. C., Sharma, M., Giles, E., & Narne, V. (2016). The Effect of Short-Term Auditory Training on Speech in Noise Perception and Cortical Auditory Evoked Potentials in Adults with Cochlear Implants. *Seminars in Hearing*, *37*(1), 84–98. https://doi.org/10.1055/s-0035-1570335

Barone, P., Chambaudie, L., Strelnikov, K., Fraysse, B., Marx, M., Belin, P., & Deguine, O. (2016). Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *83*, 259–270. https://doi.org/10.1016/j.cortex.2016.08.005

Bedoin, N. (2017). Rééquilibrer les analyses visuo-attentionnelles globales et locales pour améliorer la lecture chez des enfants dyslexiques de surface. *A.N.A.E*, *148*, 276–294.

Bedoin, N., & Medina, F. (2013). Logiciel Switchipido (Logiciel d'entraînement à l'alternance entre les traitements visuels local et global, et à l'inhibition volontaire des détails). *Www.Gnosia.Fr*.

Belfi, A. M., Evans, E., Heskje, J., Bruss, J., & Tranel, D. (2017). Musical anhedonia after focal brain damage. *Neuropsychologia*, *97*, 29–37. https://doi.org/10.1016/j.neuropsychologia.2017.01.030

Belfi, A. M., & Loui, P. (2019). Musical anhedonia and rewards of music listening: Current advances and a proposed model. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.14241

Bernstein, L. E., Eberhardt, S. P., & Auer, E. T. (2014). Audiovisual spoken word training can promote or impede auditory-only perceptual learning: Prelingually deafened adults with late-acquired cochlear implants versus normal hearing adults. *Frontiers in Psychology*, *5*, 934. https://doi.org/10.3389/fpsyg.2014.00934

Bestelmeyer, P. E. G., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsx059

Bilger, R. C. (1977). Electrical stimulation of the auditory nerve and auditory prostheses: A review of the literature. *The Annals of Otology, Rhinology & Laryngology. Supplement*, *86*(3 Pt 2 Suppl 38), 11–20. https://doi.org/10.1177/00034894770860s302

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews. Neuroscience*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 382–387. https://doi.org/10.1038/7299

Bocchio, M., Nabavi, S., & Capogna, M. (2017). Synaptic Plasticity, Engrams, and Network Oscillations in Amygdala Circuits for Storage and Retrieval of Emotional Memories. *Neuron*, *94*(4), 731–743. https://doi.org/10.1016/j.neuron.2017.03.022

Boly, M., & Seth, A. K. (2012). Modes and models in disorders of consciousness science. *Archives Italiennes De Biologie*, *150*(2–3), 172–184. https://doi.org/10.4449/aib.v150i2.1372

Borod, J. C. (1992). Interhemispheric and intrahemispheric control of emotion: A focus on unilateral brain damage. *Journal of Consulting and Clinical Psychology*, *60*(3), 339–348. https://doi.org/10.1037/0022-006X.60.3.339

Borod, J. C., Bloom, R. L., Brickman, A. M., Nakhutina, L., & Curko, E. A. (2002a). Emotional processing deficits in individuals with unilateral brain damage. *Applied Neuropsychology*, *9*(1), 23–36. https://doi.org/10.1207/S15324826AN0901_4

Borod, J. C., Bloom, R. L., Brickman, A. M., Nakhutina, L., & Curko, E. A. (2002b). Emotional Processing Deficits in Individuals With Unilateral Brain Damage. *Applied Neuropsychology*, *9*(1), 23–36. https://doi.org/10.1207/S15324826AN0901_4

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*. https://doi.org/10.1007/s10072-019-04175-w

Brancucci, A., Lugli, V., Perrucci, M. G., Del Gratta, C., & Tommasi, L. (2016). A frontal but not parietal neural correlate of auditory consciousness. *Brain Structure & Function*, *221*(1), 463–472. https://doi.org/10.1007/s00429-014-0918-2

Bregman, A. S., & McAdams, S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. *The Journal of the Acoustical Society of America*, *95*(2), 1177–1178. https://doi.org/10.1121/1.408434

Brown, S., Martinez, M. J., & Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport*, *15*(13), 2033–2037. https://doi.org/10.1097/00001756-200409150-00008

Bugannim, Y., Roth, D. A.-E., Zechoval, D., & Kishon-Rabin, L. (2019). Training of Speech Perception in Noise in Pre-Lingual Hearing Impaired Adults With Cochlear Implants Compared With Normal Hearing Adults. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *40*(3), e316–e325. https://doi.org/10.1097/MAO.0000000000002128

Butera, I. M., Stevenson, R. A., Mangus, B. D., Woynaroski, T. G., Gifford, R. H., & Wallace, M. T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. *Scientific Reports*, *8*(1), 11345. https://doi.org/10.1038/s41598-018-29598-x

Caclin, A., Bouchet, P., Djoulah, F., Pirat, E., Pernier, J., & Giard, M.-H. (2011). Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Research*, *1396*, 35–44. https://doi.org/10.1016/j.brainres.2011.04.016

Canette, L.-H., Spada, D., Pineau, M., Tillmann, B., & Bigand, E. (2016). Transfer Effects between Musical Training and Linguistic Skills in Deaf Children with Cochlear Implants. *Annals of Otolaryngology and Rhinology*, 8.

Carlson, M. L. (2020). Cochlear Implantation in Adults. *The New England Journal of Medicine*, *382*(16), 1531–1542. https://doi.org/10.1056/NEJMra1904407

Charbonneau, S., Scherzer, B. P., Aspirot, D., & Cohen, H. (2003). Perception and production of facial and prosodic emotions by chronic CVA patients. *Neuropsychologia*, *41*(5), 605–613. https://doi.org/10.1016/S0028-3932(02)00202-6

Chari, D. A., Barrett, K. C., Patel, A. D., Colgrove, T. R., Jiradejvong, P., Jacobs, L. Y., & Limb, C. J. (2020). Impact of Auditory-Motor Musical Training on Melodic Pattern Recognition in Cochlear Implant Users. *Otology & Neurotology*, *41*(4), e422. https://doi.org/10.1097/MAO.0000000000002525

Chen, J. K.-C., Chuang, A. Y. C., McMahon, C., Hsieh, J.-C., Tung, T.-H., & Li, L. P.-H. (2010). Music Training Improves Pitch Perception in Prelingually Deafened Children With Cochlear Implants. *Pediatrics*, peds.2008-3620. https://doi.org/10.1542/peds.2008-3620

Cheng, X., Liu, Y., Shu, Y., Tao, D.-D., Wang, B., Yuan, Y., Galvin, J. J., Fu, Q.-J., & Chen, B. (2018). Music Training Can Improve Music and Speech Perception in Pediatric Mandarin-Speaking Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216518759214. https://doi.org/10.1177/2331216518759214

Cheung, C. C. Y., Lee, T. M. C., Yip, J. T. H., King, K. E., & Li, L. S. W. (2006). The differential effects of thalamus and basal ganglia on facial emotion recognition. *Brain and Cognition*, *61*(3), 262–268. https://doi.org/10.1016/j.bandc.2006.01.008

Choi, J. E., Moon, I. J., Kim, E. Y., Park, H.-S., Kim, B. K., Chung, W.-H., Cho, Y.-S., Brown, C. J., & Hong, S. H. (2017). Sound Localization and Speech Perception in Noise of Pediatric Cochlear Implant Recipients: Bimodal Fitting Versus Bilateral Cochlear Implants. *Ear and Hearing*, *38*(4), 426–440. https://doi.org/10.1097/AUD.0000000000000401

Clark, C. N., Golden, H. L., & Warren, J. D. (2015). Chapter 34—Acquired amusia. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 607–631). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00034-2

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. *Psychology Press*.

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, *27*(4), 658–684. https://doi.org/10.1080/02699931.2012.732559

Cullington, H., Kitterick, P., Weal, M., & Margol-Gromada, M. (2018). Feasibility of personalised remote long-term follow-up of people with cochlear implants: A randomised controlled trial. *BMJ Open*, *8*(4), e019640. https://doi.org/10.1136/bmjopen-2017-019640

Dellacherie, D., Bigand, E., Molin, P., Baulac, M., & Samson, S. (2011). Multidimensional scaling of emotional responses to music in patients with temporal lobe resection. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *47*(9), 1107–1115. https://doi.org/10.1016/j.cortex.2011.05.007

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: Historical roots and a future incorporating "dominance." *Behavioral and Cognitive Neuroscience Reviews*, *4*(1), 3–20. https://doi.org/10.1177/1534582305276837

Deroche, M. L. D., Felezeu, M., Paquette, S., Zeitouni, A., & Lehmann, A. (2019). Neurophysiological Differences in Emotional Processing by Cochlear Implant Users, Extending Beyond the Realm of Speech. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000701

Djourno, A., & Eyries, C. (1957). [Auditory prosthesis by means of a distant electrical stimulation of the sensory nerve with the use of an indwelt coiling]. *La Presse Medicale*, *65*(63), 1417.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., & Parkin, J. L. (1978). Auditory prostheses research with multiple channel intracochlear stimulation in man. *The Annals of Otology, Rhinology, and Laryngology*, *87*(6 Pt 2), 1–39.

Edeline, J. M., & Weinberger, N. M. (1991). Thalamic short-term plasticity in the auditory system: Associative returning of receptive fields in the ventral medial geniculate body. *Behavioral Neuroscience*, *105*(5), 618–639. https://doi.org/10.1037//0735-7044.105.5.618

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2014). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology*, *5*, 1341. https://doi.org/10.3389/fpsyg.2014.01341

Eggermont, J. J. (2017). Acquired hearing loss and brain plasticity. *Hearing Research*, *343*, 176–190. https://doi.org/10.1016/j.heares.2016.05.008

Ehret, G., & Romand, R. (1997). *The central auditory system*. Oxford University Press.

Elbaum, J., & Benson, D. (Eds.). (2007). *Acquired Brain Injury: An Integrative Neuro-Rehabilitation Approach*. Springer-Verlag. https://doi.org/10.1007/978-0-387-37575-5

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810. https://doi.org/10.1002/hbm.22029

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Fakche, C., Lecaignard, F., Lévêque, Y., Fornoni, L., Tillmann, B., & Caclin, A. (2018). Impaired short-term memory for pitch in congenital amusia: A MEG/EEG study of Mismatch Negativity. *Poster Presented at MMN2018: 8th Mis-match Neg-at-iv-ity Conference, Helsinki*.

Firestone, G. M., McGuire, K., Liang, C., Zhang, N., Blankenship, C. M., Xiang, J., & Zhang, F. (2020). A Preliminary Study of the Effects of Attentive Music Listening on Cochlear Implant Users' Speech Perception, Quality of Life, and Behavioral and Objective Measures of Frequency

Change Detection. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00110

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859–869. https://doi.org/10.1016/s0896-6273(03)00669-x

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

Fourcin, A. J., Rosen, S. M., Moore, B. C., Douek, E. E., Clarke, G. P., Dodson, H., & Bannister, L. H. (1979). External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings. *British Journal of Audiology*, *13*(3), 85–107. https://doi.org/10.3109/03005367909078883

Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficits in 'tone deafness.' *Brain and Cognition*, *62*(1), 24–29. https://doi.org/10.1016/j.bandc.2006.03.005

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, *17*(9), 1442–1452. https://doi.org/10.1162/0898929054985446

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology: CB*, *19*(7), 573–576. https://doi.org/10.1016/j.cub.2009.02.058

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific Brain Networks during Explicit and Implicit Decoding of Emotional Prosody. *Cerebral Cortex*, *22*(5), 1107–1117. https://doi.org/10.1093/cercor/bhr184

Frühholz, S., & Grandjean, D. (2012). Towards a fronto-temporal neural network for the decoding of angry vocal expressions. *NeuroImage*, *62*(3), 1658–1666. https://doi.org/10.1016/j.neuroimage.2012.06.015

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *37*(1), 24–35. https://doi.org/10.1016/j.neubiorev.2012.11.002

Frühholz, S., Hofstetter, C., Cristinzio, C., Saj, A., Seeck, M., Vuilleumier, P., & Grandjean, D. (2015). Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(5), 1583–1588. https://doi.org/10.1073/pnas.1411315112

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Fu, Q.-J., & Galvin, J. J. (2007). Perceptual Learning and Auditory Training in Cochlear Implant Recipients. *Trends in Amplification*, *11*(3), 193–205. https://doi.org/10.1177/1084713807301379

Fu, Q.-J., Galvin, J. J., Wang, X., & Wu, J.-L. (2015). Benefits of Music Training in Mandarin-Speaking Pediatric Cochlear Implant Users. *Journal of Speech, Language, and Hearing Research : JSLHR*, *58*(1), 163–169. https://doi.org/10.1044/2014_JSLHR-H-14-0127

Fuller, C., Başkent, D., & Free, R. (2019). Early Deafened, Late Implanted Cochlear Implant Users Appreciate Music More Than and Identify Music as Well as Postlingual Users. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.01050

Fuller, C. D., Galvin, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of Two Music Training Approaches on Music and Speech Perception in Cochlear Implant Users. *Trends in Hearing*, *22*, 2331216518765379. https://doi.org/10.1177/2331216518765379

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.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology*, *190*(3), 597–610. https://doi.org/10.1002/cne.901900312

Galvin, 3rd JJ, Fu, Q. J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners., Melodic Contour Identification by Cochlear Implant Listeners. *Ear and Hearing, Ear*

and Hearing, *28, 28*(3, 3), 302, 302–319. https://doi.org/10.1097/01.aud.0000261689.35445.20, 10.1097/01.aud.0000261689.35445.20

Gfeller, K. (2016). Music-based training for pediatric CI recipients: A systematic analysis of published studies. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *133 Suppl 1*, S50-56. https://doi.org/10.1016/j.anorl.2016.01.010

Gfeller, Kate, Witt, S. A., Spencer, L. J., Stordahl, J., & Tomblin, B. (1999). Musical Involvement and Enjoyment of Children Who Use Cochlear Implants. *Volta Review*, *100*(4), 213–233.

Glennon, E., Svirsky, M. A., & Froemke, R. C. (2020). Auditory cortical plasticity in cochlear implant users. *Current Opinion in Neurobiology*, *60*, 108–114. https://doi.org/10.1016/j.conb.2019.11.003

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of Music Training for Perception of Emotional Speech Prosody in Deaf Children With Cochlear Implants. *Ear and Hearing*, *38*(4), 455–464. https://doi.org/10.1097/AUD.0000000000000402

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *71*, 171–182. https://doi.org/10.1016/j.cortex.2015.06.022

Gosselin, N., Peretz, I., Hasboun, D., Baulac, M., & Samson, S. (2011). Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *47*(9), 1116–1125. https://doi.org/10.1016/j.cortex.2011.05.012

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. *Neuropsychologia*, *45*(2), 236–244. https://doi.org/10.1016/j.neuropsychologia.2006.07.012

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Grasso, P. A., Làdavas, E., & Bertini, C. (2016). Compensatory Recovery after Multisensory Stimulation in Hemianopic Patients: Behavioral and Neurophysiological Components. *Frontiers in Systems Neuroscience*, *10*, 45. https://doi.org/10.3389/fnsys.2016.00045

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition*, *136*, 103614. https://doi.org/10.1016/j.bandc.2019.103614

Griffiths, T. D. (1999). *Human complex sound analysis*.

Griffiths, T. D., Warren, J. D., Dean, J. L., & Howard, D. (2004). "When the feeling's gone": A selective loss of musical emotion. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(2), 344–345. https://doi.org/10.1136/jnnp.2003.015586

Griffiths, Timothy D., & Green, G. G. R. (1999). Cortical Activation during Perception of a Rotating Wide-Field Acoustic Stimulus. *NeuroImage*, *10*(1), 84–90. https://doi.org/10.1006/nimg.1999.0464

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, *111*(3), 1340–1348.

Grondin, S. (2016). *Psychology of Perception*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31791-5

Handel, stephen. (1989). *Listening | An introduction to the perception of auditory events* (bradford book). The MIT Press. https://mitpress.mit.edu/books/listening

Harciarek, M., Heilman, K. M., & Jodzio, K. (2006). Defective comprehension of emotional faces and prosody as a result of right hemisphere stroke: Modality versus emotion-type specificity. *Journal of the International Neuropsychological Society: JINS*, *12*(6), 774–781. https://doi.org/10.1017/S1355617706061121

Hartmann, W. M. (2013). The Auditory System. In W. M. Hartmann (Ed.), *Principles of Musical Acoustics* (pp. 109–124). Springer. https://doi.org/10.1007/978-1-4614-6786-1_11

Hass, J., & Herrmann, J. M. (2012). The neural representation of time: An information-theoretic perspective. *Neural Computation*, *24*(6), 1519–1552. https://doi.org/10.1162/NECO_a_00280

Hirel, C., Lévêque, Y., Deiana, G., Richard, N., Cho, T.-H., Mechtouff, L., Derex, L., Tillmann, B., Caclin, A., & Nighoghossian, N. (2014). Acquired amusia and musical anhedonia. *Revue Neurologique*, *170*(8–9), 536–540. https://doi.org/10.1016/j.neurol.2014.03.015

Hirel, C., Nighoghossian, N., Lévêque, Y., Hannoun, S., Fornoni, L., Daligault, S., Bouchet, P., Jung, J., Tillmann, B., & Caclin, A. (2017). Verbal and musical short-term memory: Variety of auditory disorders after stroke. *Brain and Cognition*, *113*(Supplement C), 10–22. https://doi.org/10.1016/j.bandc.2017.01.003

Hoekert, M., Bais, L., Kahn, R. S., & Aleman, A. (2008). Time Course of the Involvement of the Right Anterior Superior Temporal Gyrus and the Right Fronto-Parietal Operculum in Emotional Prosody Perception. *PLoS ONE*, *3*(5). https://doi.org/10.1371/journal.pone.0002244

Hoekert, M., Vingerhoets, G., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: An rTMS study. *BMC Neuroscience*, *11*, 93. https://doi.org/10.1186/1471-2202-11-93

Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *The Journal of the Acoustical Society of America*, *120*(1), 360–374.

Hopyan, T., Manno, F. A. M., Papsin, B. C., & Gordon, K. A. (2016). Sad and happy emotion discrimination in music by children with cochlear implants. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *22*(3), 366–380. https://doi.org/10.1080/09297049.2014.992400

Huang, J., Chang, J., & Zeng, F.-G. (2017). Electro-tactile stimulation (ETS) enhances cochlear-implant Mandarin tone recognition. *World Journal of Otorhinolaryngology - Head and Neck Surgery*, *3*(4), 219–223. https://doi.org/10.1016/j.wjorl.2017.12.002

Huang, J., Lu, T., Sheffield, B., & Zeng, F.-G. (2020). Electro-Tactile Stimulation Enhances Cochlear-Implant Melody Recognition: Effects of Rhythm and Musical Training. *Ear and Hearing*, *41*(1), 106–113. https://doi.org/10.1097/AUD.0000000000000749

Huang, J., Sheffield, B., Lin, P., & Zeng, F.-G. (2017). Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise. *Scientific Reports*, *7*. https://doi.org/10.1038/s41598-017-02429-1

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. *Journal of Neuroscience*, *27*(47), 13028–13032. https://doi.org/10.1523/JNEUROSCI.3039-07.2007

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

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*(2), 632–639. https://doi.org/10.1016/j.neuropsychologia.2007.09.004

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: A two-site study. *Brain: A Journal of Neurology*, *129*(Pt 10), 2562–2570. https://doi.org/10.1093/brain/awl204

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 (New York, N.Y.: 1991)*, *21*(2), 292–299. https://doi.org/10.1093/cercor/bhq094

Innes-Brown, H., Marozeau, J., & Blamey, P. (2011). The Effect of Visual Cues on Difficulty Ratings for Segregation of Musical Streams in Listeners with Impaired Hearing. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0029327

Irvine, D. R. F. (2010, January 21). *Plasticity in the auditory pathway*. The Oxford Handbook of Auditory Science: The Auditory Brain. https://doi.org/10.1093/oxfordhb/9780199233281.013.0016

Irvine, D. R. F. (2018). Plasticity in the auditory system. *Hearing Research*, *362*, 61–73. https://doi.org/10.1016/j.heares.2017.10.011

Jafari, Z., Esmaili, M., Delbari, A., Mehrpour, M., & Mohajerani, M. H. (2017). Post-stroke acquired amusia: A comparison between right- and left-brain hemispheric damages. *NeuroRehabilitation*, *40*(2), 233–241. https://doi.org/10.3233/NRE-161408

Janata, P., Birk, J. L., Horn, J. D. V., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The Cortical Topography of Tonal Structures Underlying Western Music. *Science*, *298*(5601), 2167–2170. https://doi.org/10.1126/science.1076262

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 121–140. https://doi.org/10.3758/CABN.2.2.121

Jiam, N. T., & Limb, C. J. (2019). Rhythm processing in cochlear implant−mediated music perception. *Annals of the New York Academy of Sciences*, *1453*(1), 22–28. https://doi.org/10.1111/nyas.14130

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*(9), 2630–2639. https://doi.org/10.1016/j.neuropsychologia.2010.05.009

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition*, *40*(7), 1109–1121. https://doi.org/10.3758/s13421-012-0208-2

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163. https://doi.org/10.1093/brain/123.1.155

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*(2), 164–170. https://doi.org/10.1016/S0959-4388(99)80022-1

Kalathottukaren, R. T., Purdy, S. C., & Ballard, E. (2015). Prosody perception and musical pitch discrimination in adults using cochlear implants. *International Journal of Audiology*, *54*(7), 444–452. https://doi.org/10.3109/14992027.2014.997314

Kéïta, L., Bedoin, N., Burack, J. A., & Lepore, F. (2014). Switching between global and local levels: The level repetition effect and its hemispheric asymmetry. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00252

Khalfa, Delbe, C., Bigand, E., Reynaud, E., Chauvel, P., & Liéégeois-Chauvel, C. (2007). Positive and Negative Music Recognition Reveals a Specialization of Mesio-Temporal Structures in Epileptic Patients. *Music Perception: An Interdisciplinary Journal*, *25*(4), 295–302. https://doi.org/10.1525/mp.2008.25.4.295

Khalfa, Guye, M., Peretz, I., Chapon, F., Girard, N., Chauvel, P., & Liégeois-Chauvel, C. (2008). Evidence of lateralized anteromedial temporal structures involvement in musical emotion processing. *Neuropsychologia*, *46*(10), 2485–2493. https://doi.org/10.1016/j.neuropsychologia.2008.04.009

Kim, W. B., & Cho, J.-H. (2017). Encoding of Discriminative Fear Memory by Input-Specific LTP in the Amygdala. *Neuron*, *95*(5), 1129-1146.e5. https://doi.org/10.1016/j.neuron.2017.08.004

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, *17*(5), 307–321. https://doi.org/10.1038/nrn.2016.22

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Koelsch, S. (2015). Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Annals of the New York Academy of Sciences*, *1337*(1), 193–201. https://doi.org/10.1111/nyas.12684

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

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J.-D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping*, *34*(8), 1971–1981. https://doi.org/10.1002/hbm.22041

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kral, A., & Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends in Neurosciences*, *35*(2), 111–122. https://doi.org/10.1016/j.tins.2011.09.004

Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia*, *41*(8), 1082–1090.

Kumar, S., Kriegstein, K. von, Friston, K., & Griffiths, T. D. (2012). Features versus Feelings: Dissociable Representations of the Acoustic Features and Valence of Aversive Sounds. *Journal of Neuroscience*, *32*(41), 14184–14192. https://doi.org/10.1523/JNEUROSCI.1759-12.2012

Langers, D. R. M., Backes, W. H., & Dijk, P. van. (2007). Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage*, *34*(1), 264–273. https://doi.org/10.1016/j.neuroimage.2006.09.002

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. *Auditory Cognitive Neuroscience*, 343. https://doi.org/10.3389/fnins.2015.00343

Lerousseau, J. P., Hidalgo, C., & Schön, D. (2020). Musical Training for Auditory Rehabilitation in Hearing Loss. *Journal of Clinical Medicine*, *9*(4), 1058. https://doi.org/10.3390/jcm9041058

Leveque, Y., Fauvel, B., Groussard, M., Caclin, A., Albouy, P., Platel, H., & Tillmann, B. (2016). Altered intrinsic connectivity of the auditory cortex in congenital amusia. *Journal of Neurophysiology*, jn.00663.2015. https://doi.org/10.1152/jn.00663.2015

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*. https://doi.org/10.1037/neu0000461

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The Language, Tone and Prosody of Emotions: Neural Substrates and Dynamics of Spoken-Word Emotion Perception. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00506

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*. https://doi.org/10.1038/srep34911

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion (Washington, D.C.)*, *11*(5), 1021–1031. https://doi.org/10.1037/a0024521

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., & Wong, P. C. M. (2017). Perceptual learning of pitch direction in congenital amusia: Evidence from Chinese speakers. *Music Perception*, *34*(3), 335–351.

Liu, F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia*, *66*, 111–118. https://doi.org/10.1016/j.neuropsychologia.2014.11.001

Lo, C. Y., McMahon, C. M., Looi, V., & Thompson, W. F. (2015). Melodic Contour Training and Its Effect on Speech in Noise, Consonant Discrimination, and Prosody Perception for Cochlear Implant Recipients. *Behavioural Neurology*, *2015*, 352869. https://doi.org/10.1155/2015/352869

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology*, *64*, 109–143. https://doi.org/10.1159/000094648

Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: Results from congenital amusia. *Frontiers in Psychology*, *6*, 1340. https://doi.org/10.3389/fpsyg.2015.01340

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(33), 10215–10220. https://doi.org/10.1523/JNEUROSCI.1701-09.2009

Luo, X., & Warner, B. (2020). Effect of instrument timbre on musical emotion recognition in normal-hearing listeners and cochlear implant users. *The Journal of the Acoustical Society of America*, *147*(6), EL535–EL539. https://doi.org/10.1121/10.0001475

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: An auditory-motor feedback disorder? *Restorative Neurology and Neuroscience*, *25*(3–4), 323–334.

Marin, M. M., Thompson, W. F., Gingras, B., & Stewart, L. (2015). Affective evaluation of simultaneous tone combinations in congenital amusia. *Neuropsychologia*, *78*, 207–220. https://doi.org/10.1016/j.neuropsychologia.2015.10.004

Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., & Deguine, O. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. *Ear and Hearing*, *36*(2), 239–248. https://doi.org/10.1097/AUD.0000000000000105

McAdams, S. (2015). *Perception et cognition de la musique*. http://www.vrin.fr/book.php?code=9782711625772

McAdams, S., & Bigand, E. (1993). Introduction à la cognition auditive. In *Penser les Sons: Psychologie Cognitive de l'Audition* (Oxford University Press). http://articles.ircam.fr/textes/McAdams93b/

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification*, *8*(2), 49–82. https://doi.org/10.1177/108471380400800203

Mcdonald, C., & Stewart, L. (2008). Uses and functions of music in congenital amusia. *Music Perception 25(4):345*. https://doi.org/10.1525/MP.2008.25.4.345

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Milesi, V., Cekic, S., Péron, J., Frühholz, S., Cristinzio, C., Seeck, M., & Grandjean, D. (2014). Multimodal emotion perception after anterior temporal lobectomy (ATL). *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00275

Moore, B. C. (2003). *An Introduction to the Psychology of Hearing* (Academic Press, San Diego, USA). https://www.abebooks.com/9780125056281/Introduction-Psychology-Hearing-Moore-B.C.J-0125056281/plp

Moulin, A., & Garcia, S. (2012). Test Audimots. *Agence Prot Programme*.

Moulin, A., & Richard, C. (2015). Lexical Influences on Spoken Spondaic Word Recognition in Hearing-Impaired Patients. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00476

Moulin, A., & Richard, C. (2016). Sources of variability of speech, spatial, and qualities of hearing scale (SSQ) scores in normal-hearing and hearing-impaired populations. *International Journal of Audiology*, *55*(2), 101–109. https://doi.org/10.3109/14992027.2015.1104734

Moulin, A., Vergne, J., Gallego, S., & Micheyl, C. (2019). A New Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster, and Comparative Analyses. *Ear and Hearing*, *40*(4), 938–950. https://doi.org/10.1097/AUD.0000000000000675

Nan, Y., Huang, W. T., Wang, W. J., Liu, C., & Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (MMN) among Mandarin speakers with congenital amusia. *Biological Psychology*, *113*, 59–67. https://doi.org/10.1016/j.biopsycho.2015.11.010

Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Research*, *405*(2), 268–283. https://doi.org/10.1016/0006-8993(87)90296-4

Neville, H. J., Schmidt, A., & Kutas, M. (1983). Altered visual-evoked potentials in congenitally deaf adults. *Brain Research*, *266*(1), 127–132. https://doi.org/10.1016/0006-8993(83)91314-8

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 490–493. https://doi.org/10.1111/j.1749-6632.2009.04855.x

Nolden, S., Rigoulot, S., Jolicoeur, P., & Armony, J. L. (2017). Effects of musical expertise on oscillatory brain activity in response to emotional sounds. *Neuropsychologia*, *103*, 96–105. https://doi.org/10.1016/j.neuropsychologia.2017.07.014

Nordström, H., & Laukka, P. (2019). The time course of emotion recognition in speech and music. *The Journal of the Acoustical Society of America*, *145*(5), 3058. https://doi.org/10.1121/1.5108601

Omigie, D. (2016). Basic, specific, mechanistic? Conceptualizing musical emotions in the brain. *Journal of Comparative Neurology*, *524*(8), 1676–1686. https://doi.org/10.1002/cne.23854

Omigie, D., Dellacherie, D., Hasboun, D., Clément, S., Baulac, M., Adam, C., & Samson, S. (2015). Intracranial markers of emotional valence processing and judgments in music. *Cognitive Neuroscience*, *6*(1), 16–23. https://doi.org/10.1080/17588928.2014.988131

Omigie, D., Dellacherie, D., Hasboun, D., George, N., Clement, S., Baulac, M., Adam, C., & Samson, S. (2015). An Intracranial EEG Study of the Neural Dynamics of Musical Valence Processing. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(11), 4038–4047. https://doi.org/10.1093/cercor/bhu118

Omigie, D., Müllensiefen, D., & Stewart, L. (2012). The Experience of Music in Congenital Amusia. *Music Perception: An Interdisciplinary Journal*, *30*(1), 1–18. https://doi.org/10.1525/mp.2012.30.1.1

Pak, C. L., & Katz, W. F. (2019). Recognition of emotional prosody by Mandarin-speaking adults with cochlear implants. *The Journal of the Acoustical Society of America*, *146*(2), EL165–EL171. https://doi.org/10.1121/1.5122192

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, *94*(1), 26–40. https://doi.org/10.1016/0013-4694(94)00209-4

Pantev, Christo, & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2140–2154. https://doi.org/10.1016/j.neubiorev.2011.06.010

Pantev, Christo, Paraskevopoulos, E., Kuchenbuch, A., Lu, Y., & Herholz, S. C. (2015). Musical expertise is related to neuroplastic changes of multisensory nature within the auditory cortex. *The European Journal of Neuroscience*, *41*(5), 709–717. https://doi.org/10.1111/ejn.12788

Paquette, S., Ahmed, G. D., Goffi-Gomez, M. V., Hoshino, A. C. H., Peretz, I., & Lehmann, A. (2018). Musical and vocal emotion perception for cochlear implants users. *Hearing Research*. https://doi.org/10.1016/j.heares.2018.08.009

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The Plastic Human Brain Cortex. *Annual Review of Neuroscience*, *28*(1), 377–401. https://doi.org/10.1146/annurev.neuro.27.070203.144216

Passamonti, C., Bertini, C., & Làdavas, E. (2009). Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia*, *47*(2), 546–555. https://doi.org/10.1016/j.neuropsychologia.2008.10.008

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*(3), 310–313. https://doi.org/10.1016/j.bandc.2004.10.003

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech Intonation Perception Deficits in Musical Tone Deafness (congenital Amusia). *Music Perception: An Interdisciplinary Journal*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

Pattisapu, P., Lindquist, N. R., Appelbaum, E. N., Silva, R. C., Vrabec, J. T., & Sweeney, A. D. (2020). A Systematic Review of Cochlear Implant Outcomes in Prelingually-deafened, Late-implanted Patients. *Otology & Neurotology*, *Publish Ahead of Print*. https://doi.org/10.1097/MAO.0000000000002555

Paulmann, S., Pell, M. D., & Kotz, S. A. (2005). Emotional prosody recognition in BG-patients: Disgust recognition revisited. *Brain and Language*, *95*(1), 143–144.

Paulmann, Silke, Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research*, *1217*, 171–178. https://doi.org/10.1016/j.brainres.2008.04.032

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 275–288. https://doi.org/10.3758/CABN.3.4.275

Peng, S.-C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing*, *29*(3), 336–351. https://doi.org/10.1097/AUD.0b013e318168d94d

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*, 58–75.

Peretz, I., Cummings, S., & Dubé, M.-P. (2007). The Genetics of Congenital Amusia (Tone Deafness): A Family-Aggregation Study. *American Journal of Human Genetics*, *81*(3), 582–588.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. https://doi.org/10.1016/S0010-0277(98)00043-2

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics: EJHG*, *25*(5), 625–630. https://doi.org/10.1038/ejhg.2017.15

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, B., Andersen, A. S. F., Haumann, N. T., Højlund, A., Dietz, M. J., Michel, F., Riis, S. K., Brattico, E., & Vuust, P. (2020). The CI MuMuFe – A New MMN Paradigm for Measuring Music Discrimination in Electric Hearing. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00002

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch--time interference in congenital amusia: Evidence from an implicit test. *Attention, Perception & Psychophysics*, *72*(3), 763–774. https://doi.org/10.3758/APP.72.3.763

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (2005). *Pitch: Neural Coding and Perception*. Springer.

Pralus, A., Belfi, A., Hirel, C., Lévêque, Y., Fornoni, L., Bigand, E., Jung, J., Tranel, D., Nighoghossian, N., Tillmann, B., & Caclin, A. (2020). Recognition of musical emotions and their perceived intensity after unilateral brain damage. *Cortex*. https://doi.org/10.1016/j.cortex.2020.05.015

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

Pralus, Agathe, Belfi, A., Hirel, C., Lévêque, Y., Fornoni, L., Bigand, E., Jung, J., Tranel, D., Nighoghossian, N., Tillmann, B., & Caclin, A. (2020). Recognition of musical emotions and their perceived intensity after unilateral brain damage. *Cortex*, *130*, 78–93. https://doi.org/10.1016/j.cortex.2020.05.015

Pralus, Agathe, Lévêque, Y., Blain, S., Bouchet, P., Tillmann, B., & Caclin, A. (2019). Training pitch perception in cochlear implant users via audiovisual integration: A pilot study in normal-hearing controls. *Implants&Music Congress, Montréal 2019*, *Poster presentation*.

Rauschecker, J. P. (2018). Where, When, and How: Are they all Sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *98*, 262–268. https://doi.org/10.1016/j.cortex.2017.10.020

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., Pons, T., & Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *Journal of Comparative Neurology*, *382*(1), 89–103. https://doi.org/10.1002/(SICI)1096-9861(19970526)382:1<89::AID-CNE6>3.0.CO;2-G

Rayes, H., Al-Malky, G., & Vickers, D. (2019). Systematic Review of Auditory Training in Pediatric Cochlear Implant Recipients. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(5), 1574–1593. https://doi.org/10.1044/2019_JSLHR-H-18-0252

Riley, P. E., Ruhl, D. S., Camacho, M., & Tolisano, A. M. (2018). Music Appreciation after Cochlear Implantation in Adult Patients: A Systematic Review. *Otolaryngology–Head and Neck Surgery*, *158*(6), 1002–1010. https://doi.org/10.1177/0194599818760559

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, *6*(4), 288–304. https://doi.org/10.1006/nimg.1997.0304

Rocca, C. (2012). A Different Musical Perspective: Improving Outcomes in Music through Habilitation, Education, and Training for Children with Cochlear Implants. *Seminars in Hearing*, *33*(04), 425–433. https://doi.org/10.1055/s-0032-1329229

Rochette, F., & Bigand, E. (2009). Long-term Effects of Auditory Training in Severely or Profoundly Deaf Children. *Annals of the New York Academy of Sciences*, *1169*(1), 195–198. https://doi.org/10.1111/j.1749-6632.2009.04793.x

Rochette, F., Moussard, A., & Bigand, E. (2014). Music lessons improve auditory perceptual and cognitive performance in deaf children. *Frontiers in Human Neuroscience*, *8*, 488. https://doi.org/10.3389/fnhum.2014.00488

Roman, S., Rochette, F., Triglia, J.-M., Schön, D., & Bigand, E. (2016). Auditory training improves auditory performance in cochlear implanted children. *Hearing Research*, *337*, 89–95. https://doi.org/10.1016/j.heares.2016.05.003

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research*, *1188*, 87–99. https://doi.org/10.1016/j.brainres.2007.10.049

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7295–7300. https://doi.org/10.1073/pnas.0609419104

Sackeim, H. A., Greenberg, M. S., Weiman, A. L., Gur, R. C., Hungerbuhler, J. P., & Geschwind, N. (1982). Hemispheric Asymmetry in the Expression of Positive and Negative Emotions: Neurologic Evidence. *Archives of Neurology*, *39*(4), 210–218. https://doi.org/10.1001/archneur.1982.00510160016003

Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 42–52. https://doi.org/10.1016/j.heares.2013.07.016

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, *14*(2), 257–262. https://doi.org/10.1038/nn.2726

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and Ventral Pathways for Prosody. *Current Biology: CB*, *25*(23), 3079–3085. https://doi.org/10.1016/j.cub.2015.10.009

Samson, S., Zatorre, R. J., & Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain: A Journal of Neurology*, *125*(Pt 3), 511–523. https://doi.org/10.1093/brain/awf051

Sander, K., & Scheich, H. (2005). Left Auditory Cortex and Amygdala, but Right Insula Dominance for Human Laughing and Crying. *Journal of Cognitive Neuroscience*, *17*(10), 1519–1531. https://doi.org/10.1162/089892905774597227

Sandmann, P., Plotz, K., Hauthal, N., de Vos, M., Schönfeld, R., & Debener, S. (2015). Rapid bilateral improvement in auditory cortex activity in postlingually deafened adults following cochlear implantation. *Clinical Neurophysiology*, *126*(3), 594–607. https://doi.org/10.1016/j.clinph.2014.06.029

Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., & Hietanen, M. (2009). Cognitive deficits associated with acquired amusia after stroke: A neuropsychological follow-up study. *Neuropsychologia*, *47*(12), 2642–2651. https://doi.org/10.1016/j.neuropsychologia.2009.05.015

Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., Hietanen, M., & Pihko, E. (2010). Auditory and Cognitive Deficits Associated with Acquired Amusia after Stroke: A Magnetoencephalography and Neuropsychological Follow-Up Study. *PLoS ONE*, *5*(12). https://doi.org/10.1371/journal.pone.0015157

Sato, T., Yabushita, T., Sakamoto, S., Katori, Y., & Kawase, T. (2019). In-home auditory training using audiovisual stimuli on a tablet computer: Feasibility and preliminary results. *Auris Nasus Larynx*. https://doi.org/10.1016/j.anl.2019.09.006

Satoh, M., Kato, N., Tabei, K.-I., Nakano, C., Abe, M., Fujita, R., Kida, H., Tomimoto, H., & Kondo, K. (2016). A case of musical anhedonia due to right putaminal hemorrhage: A disconnection syndrome between the auditory cortex and insula. *Neurocase*, *22*(6), 518–525. https://doi.org/10.1080/13554794.2016.1264609

Satoh, M., Nakase, T., Nagata, K., & Tomimoto, H. (2011). Musical anhedonia: Selective loss of emotional experience in listening to music. *Neurocase*, *17*(5), 410–417. https://doi.org/10.1080/13554794.2010.532139

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schuchard, J., Nerantzini, M., & Thompson, C. K. (2017). Implicit learning and implicit treatment outcomes in individuals with aphasia. *Aphasiology*, *31*(1), 25–48. https://doi.org/10.1080/02687038.2016.1147526

Schuchard, J., & Thompson, C. K. (2014). Implicit and explicit learning in individuals with agrammatic aphasia. *Journal of Psycholinguistic Research*, *43*(3), 209–224. https://doi.org/10.1007/s10936-013-9248-4

Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*(1), 189–196. https://doi.org/10.1111/j.1460-9568.2010.07470.x

Schulze, Katrin, 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

Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain: A Journal of Neurology*, *123 Pt 3*, 546–559.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Sharp, A., Delcenserie, A., & Champoux, F. (2018). Auditory Event-Related Potentials Associated With Music Perception in Cochlear Implant Users. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00538

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology*, *11*(4), 505–509. https://doi.org/10.1016/S0959-4388(00)00241-5

Shirvani, S., Jafari, Z., Motasaddi Zarandi, M., Jalaie, S., Mohagheghi, H., & Tale, M. R. (2016). Emotional Perception of Music in Children With Bimodal Fitting and Unilateral Cochlear Implant. *The Annals of Otology, Rhinology, and Laryngology*, *125*(6), 470–477. https://doi.org/10.1177/0003489415619943

Shirvani, S., Jafari, Z., Sheibanizadeh, A., Motasaddi Zarandy, M., & Jalaie, S. (2014). Emotional perception of music in children with unilateral cochlear implants. *Iranian Journal of Otorhinolaryngology*, *26*(77), 225–233.

Sihvonen, A. J., Ripollés, P., Leo, V., Rodríguez-Fornells, A., Soinila, S., & Särkämö, T. (2016). Neural Basis of Acquired Amusia and Its Recovery after Stroke. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(34), 8872–8881. https://doi.org/10.1523/JNEUROSCI.0709-16.2016

Sihvonen, A. J., Ripollés, P., Rodríguez-Fornells, A., Soinila, S., & Särkämö, T. (2017). Revisiting the Neural Basis of Acquired Amusia: Lesion Patterns and Structural Changes Underlying Amusia Recovery. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00426

Sihvonen, A. J., Ripollés, P., Särkämö, T., Leo, V., Rodríguez-Fornells, A., Saunavaara, J., Parkkola, R., & Soinila, S. (2017). Tracting the neural basis of music: Deficient structural connectivity underlying acquired amusia. *Cortex*, *97*, 255–273. https://doi.org/10.1016/j.cortex.2017.09.028

Sihvonen, A. J., Särkämö, T., Ripollés, P., Leo, V., Saunavaara, J., Parkkola, R., Rodríguez-Fornells, A., & Soinila, S. (2017). Functional neural changes associated with acquired amusia across different stages of recovery after stroke. *Scientific Reports*, *7*(1), 11390. https://doi.org/10.1038/s41598-017-11841-6

Simmons, F. B. (1966). Electrical stimulation of the auditory nerve in man. *Archives of Otolaryngology (Chicago, Ill.: 1960)*, *84*(1), 2–54. https://doi.org/10.1001/archotol.1966.00760030004003

Slee, S. J., & David, S. V. (2015). Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in the Auditory Midbrain. *Journal of Neuroscience*, *35*(38), 13090–13102. https://doi.org/10.1523/JNEUROSCI.1671-15.2015

Smith, L., Bartel, L., Joglekar, S., & Chen, J. (2017). Musical Rehabilitation in Adult Cochlear Implant Recipients With a Self-administered Software. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *38*(8), e262–e267. https://doi.org/10.1097/MAO.0000000000001447

Snow, W. B. (1936). Change of Pitch with Loudness at Low Frequencies. *The Journal of the Acoustical Society of America*, *8*(1), 14–19. https://doi.org/10.1121/1.1915846

Spalding, K. L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H. B., Boström, E., Westerlund, I., Vial, C., Buchholz, B. A., Possnert, G., Mash, D. C., Druid, H., & Frisén, J. (2013). Dynamics of hippocampal neurogenesis in adult humans. *Cell*, *153*(6), 1219–1227. https://doi.org/10.1016/j.cell.2013.05.002

Spangmose, S., Hjortkjær, J., & Marozeau, J. (2019). Perception of Musical Tension in Cochlear Implant Listeners. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/fnins.2019.00987

Stewart, L. (2011). Characterizing congenital amusia. *Quarterly Journal of Experimental Psychology (2006)*, *64*(4), 625–638. https://doi.org/10.1080/17470218.2011.552730

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain: A Journal of Neurology*, *129*(Pt 10), 2533–2553. https://doi.org/10.1093/brain/awl171

Strelnikov, K., Rouger, J., Barone, P., & Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scandinavian Journal of Psychology*, *50*(5), 437–444. https://doi.org/10.1111/j.1467-9450.2009.00741.x

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Démonet, J.-F., Déguine, O., & Barone, P. (2015). Increased audiovisual integration in cochlear-implanted deaf patients: Independent components analysis of longitudinal positron emission tomography data. *The European Journal of Neuroscience*, *41*(5), 677–685. https://doi.org/10.1111/ejn.12827

Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Reviews Neuroscience*, *4*(10), 783–794. https://doi.org/10.1038/nrn1222

Sun, Y., Lu, X., Ho, H. T., Johnson, B. W., Sammler, D., & Thompson, W. F. (2018). Syntactic processing in music and language: Parallel abnormalities observed in congenital amusia. *NeuroImage. Clinical*, *19*, 640–651. https://doi.org/10.1016/j.nicl.2018.05.032

Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: Distinct cerebral responses during perception of different laughter types. *NeuroImage*, *53*(4), 1264–1271. https://doi.org/10.1016/j.neuroimage.2010.06.028

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, *12*(10), e0186773. https://doi.org/10.1371/journal.pone.0186773

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex11Portions of this work were presented at the annual meeting of the International Society for Magnetic Resonance in Medicine (1996) and the annual meeting of the Association for Research in Otolaryngology (1997). *Hearing Research*, *150*(1), 225–244. https://doi.org/10.1016/S0378-5955(00)00203-3

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity. *Journal of Neurophysiology*, *91*(3), 1282–1296. https://doi.org/10.1152/jn.01125.2002

Thompson, W. F., Balkwill, L. L., & Vernescu, R. (2000). Expectancies generated by recent exposure to melodic sequences. *Memory & Cognition*, *28*(4), 547–555. https://doi.org/10.3758/bf03201245

Thompson, William Forde, Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Thompson, William Forde, Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., & von Cramon, D. Y. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, *31*(4), 1771–1782. https://doi.org/10.1016/j.neuroimage.2006.02.028

Tillmann, Barbara, Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handbook of Clinical Neurology*, *129*, 589–605. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, Barbara, Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital Amusia (or Tone-Deafness) Interferes with Pitch Processing in Tone Languages. *Frontiers in Psychology*, *2*, 120. https://doi.org/10.3389/fpsyg.2011.00120

Tillmann, Barbara, Hirel, C., Lévêque, Y., & Caclin, A. (2017). Musical connections: Music perception and neurological deficits. *The Routledge Companion to Music Cognition*, *chapter 19*.

Tillmann, Barbara, Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*(2), 145–161. https://doi.org/10.1016/S0926-6410(02)00245-8

Tillmann, Barbara, Lalitte, P., Albouy, P., Caclin, A., & Bigand, E. (2016). Discrimination of tonal and atonal music in congenital amusia: The advantage of implicit tasks. *Neuropsychologia*, *85*, 10–18. https://doi.org/10.1016/j.neuropsychologia.2016.02.027

Tillmann, Barbara, Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640, Part B*, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

Tillmann, Barbara, Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., & Peretz, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *The Journal of the Acoustical Society of America*, *130*(6), 4089–4096. https://doi.org/10.1121/1.3658447

Tillmann, Barbara, 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

Toni, N., Teng, E. M., Bushong, E. A., Aimone, J. B., Zhao, C., Consiglio, A., van Praag, H., Martone, M. E., Ellisman, M. H., & Gage, F. H. (2007). Synapse formation on neurons born in the adult hippocampus. *Nature Neuroscience*, *10*(6), 727–734. https://doi.org/10.1038/nn1908

Torppa, R., Faulkner, A., Huotilainen, M., Jaervikivi, J., Lipsanen, J., Laasonen, M., & Vainio, M. (2014). The perception of prosody and associated auditory cues in early-implanted children:

The role of auditory working memory and musical activities. *INTERNATIONAL JOURNAL OF AUDIOLOGY*, *53*(3), 182–191.

Torppa, Ritva, & Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hearing Research*, *380*, 108–122. https://doi.org/10.1016/j.heares.2019.06.003

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping Aesthetic Musical Emotions in the Brain. *Cerebral Cortex*, *22*(12), 2769–2783. https://doi.org/10.1093/cercor/bhr353

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55–64. https://doi.org/10.1016/j.neuroimage.2014.09.009

Turrigiano, G. (2011). Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement. *Annual Review of Neuroscience*, *34*(1), 89–103. https://doi.org/10.1146/annurev-neuro-060909-153238

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*.

Vandali, A., Sly, D., Cowan, R., & van Hoesel, R. (2015). Training of cochlear implant users to improve pitch perception in the presence of competing place cues. *Ear and Hearing*, *36*(2), e1–e13. https://doi.org/10.1097/AUD.0000000000000109

Vuilleumier, P., & Trost, W. (2015). Music and emotions: From enchantment to entrainment. *Annals of the New York Academy of Sciences*, *1337*, 212–222. https://doi.org/10.1111/nyas.12676

Waddington, E., Jaekel, B. N., Tinnemore, A. R., Gordon-Salant, S., & Goupell, M. J. (2020). Recognition of Accented Speech by Cochlear-Implant Listeners: Benefit of Audiovisual Cues. *Ear and Hearing*, *Publish Ahead of Print*. https://doi.org/10.1097/AUD.0000000000000842

Wallace, M. N., Johnston, P. W., & Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research*, *143*(4), 499–508. https://doi.org/10.1007/s00221-002-1014-z

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, *5*(4), 279–290. https://doi.org/10.1038/nrn1366

Weinberger, N. M. (2010). *The cognitive auditory cortex: Vol. The Oxford Handbook of Auditory Science: the Auditory Brain* (A. Palmer, A. Rees). Oxford University Press.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. *Cerebral Cortex*, *14*(12), 1384–1389. https://doi.org/10.1093/cercor/bhh099

Wildgruber, Dirk, Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277–281. https://doi.org/10.1080/17549500902943043

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory (Hove, England)*, *18*(6), 657–669. https://doi.org/10.1080/09658211.2010.501339

Yucel, E., Sennaroglu, G., & Belgin, E. (2009). The family oriented musical training for children with cochlear implants: Speech and musical perception results of two year follow-up. *International Journal of Pediatric Otorhinolaryngology*, *73*(7), 1043–1052. https://doi.org/10.1016/j.ijporl.2009.04.009

Yuvaraj, R., Murugappan, M., Norlinah, M. I., Sundaraj, K., & Khairiyah, M. (2013). Review of Emotion Recognition in Stroke Patients. *Dementia and Geriatric Cognitive Disorders*, *36*(3–4), 179–196. https://doi.org/10.1159/000353440

Zald, D. H., & Pardo, J. V. (2002). The Neural Correlates of Aversive Auditory Stimulation. *NeuroImage*, *16*(3, Part A), 746–753. https://doi.org/10.1006/nimg.2002.1115

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(4), 1908–1919.

Zatorre, Robert 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

Zhou, Q., Gu, X., & Liu, B. (2019). [The music quality feeling and music perception of adult cochlear implant recipients]. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi = Journal of Clinical*

Otorhinolaryngology, Head, and Neck Surgery, *33*(1), 47–51. https://doi.org/10.13201/j.issn.1001-1781.2019.01.010