



HAL
open science

Memory for random time patterns in the sensory system

Hijee Kang

► **To cite this version:**

Hijee Kang. Memory for random time patterns in the sensory system. Psychology. Université Paris sciences et lettres, 2017. English. NNT : 2017PSLEE079 . tel-04368436

HAL Id: tel-04368436

<https://theses.hal.science/tel-04368436>

Submitted on 1 Jan 2024

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

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

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Normale Supérieure

Memory for random time patterns in sensory systems

Ecole doctorale n°158

Cerveau, Cognition, comportement

Spécialité Sciences Cognitives

Soutenue par HiJee KANG
le 18 Décembre 2017

Dirigée par **Daniel PRESSNITZER**

COMPOSITION DU JURY :

Mme. CHAIT Maria
UCL, Rapporteur

M. DEMANY Laurent
Université de Bordeaux, Rapporteur

M. BOURDIEU Laurent
École Normale Supérieure, Président du jury

M. SCHNUPP Jan
City University of Hong Kong, Membre du jury

M. THORP Simon
Université de Toulouse, Membre du jury

Résumé

Le temps est une dimension universelle traitée par les systèmes sensoriels, qui est essentielle pour attribuer un sens à des stimuli comme la parole ou la musique pour l'audition. Cependant, les mécanismes requis pour le traitement temporel restent en grande part méconnus. Dans cette thèse, nous avons examiné un type de mécanisme faisant sans doute partie intégrante de tout traitement temporel: la formation de nouvelles traces mnésiques pour l'information temporelle. Nous avons étudié principalement la modalité auditive, mais aussi d'autres modalités sensorielles, comme le toucher, la vision, et la stimulation électrique directe du système auditif périphérique avec un implant cochléaire. Toutes les expériences ont utilisé un nouveau paradigme expérimental, adapté de précédentes études conçues pour étudier la mémoire auditive (Agus, Thorpe, & Pressnitzer, 2010). Au lieu d'utiliser du bruit comme stimulus, nous avons utilisé des séquences d'intervalles de temps irréguliers délimités par de brèves impulsions d'énergie, adaptées à la modalité étudiée. Dans une première série d'expériences, nous avons étudié la modalité auditive chez des auditeurs normo-entendants, en utilisant des trains de clics audio comme stimuli. Nous avons démontré, pour la première fois, un apprentissage rapide de sons contenant uniquement des informations temporelles. Dans une seconde série d'expériences, nous avons appliqué le même paradigme à trois modalités sensorielles (audition, toucher et vision), en utilisant des clics audio, des impulsions de mouvement au bout des doigts, et des flashes de lumière pour délimiter les intervalles de temps dans les différences modalités. Nous avons observé des formes qualitativement similaires d'apprentissage perceptif pour les trois modalités, avec un apprentissage rapide dans tous les cas, ainsi qu'un transfert d'apprentissage au toucher ou à la vision pour des séquences initialement apprises de façon auditive. Dans une troisième série d'expériences, nous avons testé des malentendants stimulés électriquement par leur implant cochléaire avec des séquences d'impulsions irrégulières. Nous avons trouvé des indications d'une plasticité préservée pour l'apprentissage rapide des informations temporelles chez ces auditeurs. Enfin, nous présentons des résultats préliminaires en utilisant une nouvelle technique susceptible de révéler certains des mécanismes neuronaux sous-jacents à l'apprentissage perceptuel rapide. Nous avons mesuré la dilatation pupillaire pendant que les auditeurs effectuaient la tâche de mémoire auditive et observé des changements systématiques de la taille de la pupille avec l'apprentissage. En conclusion, la thèse montre une capacité remarquable des systèmes perceptifs à apprendre des séquences temporelles complexes lorsqu'elles apparaissent plusieurs fois dans l'environnement, et suggère de nouvelles méthodes expérimentales pour étudier plus avant les mécanismes neuronaux sous-jacents.

Mots Clés

audition, apprentissage perceptif, traitement temporelle, vision, toucher, activité neuronal

Abstract

Time is a universal feature of all information processed by sensory systems, and temporal patterning is often essential for attributing meaning to external stimuli such as speech or music in audition. However, many of the mechanisms needed for temporal processing are still unclear. In this thesis, we investigated one type of mechanism arguably integral to any kind of temporal processing: the formation of novel memories for temporal patterns. We studied mainly the auditory modality, but also other sensory modalities such touch, vision, and electric hearing with a cochlear implant. All experiments used a novel experimental paradigm, adapted from a previous study designed to explore auditory memory of random noise (Agus, Thorpe, & Pressnitzer, 2010). Instead of using noise as the complex stimulus to learn, we used irregular time patterns made of random time intervals delineated by modality-adapted brief energy pulses. In a first series of experiments, we investigated the auditory modality in normal hearing listeners, using click trains as stimuli. We demonstrated for the first time a rapid learning of stimuli containing solely temporal cues. In a second series of experiments, we applied the same paradigm to multiple sensory modalities (audition, touch, and vision), using audio clicks, motion pulses to the fingertips, and light to delineate time intervals. We found a qualitatively similar forms of perceptual learning for all three modalities, with rapid learning in all cases, as well as a transfer of learning to touch or vision for patterns learnt initially learnt in audition. In a third series of experiments, we tested hearing impaired listeners stimulated through their cochlear implant with sequences of electrical pulses. We found evidence for preserved plasticity for the rapid learning of time patterns in those listeners. Finally, we present preliminary data using a novel technique for studying the underlying neural mechanisms of rapid perceptual learning. We measured pupil dilation while listeners performed the memory task and observed systematic changes in pupil size with perceptual learning. In conclusion, the thesis shows a remarkable ability of perceptual systems to learn complex time patterns as they re-occur in the environment, and suggests new experimental methods to further study the underlying neural mechanisms.

Keywords

audition, perceptual learning, temporal processing, vision, touch, neuronal activity



Thèse de doctorat de l'École Normale Supérieure
École Doctorale Cerveau, Comportement, et Cognition - ED3C n° 158

Présentée par
HiJee Kang

Pour obtenir le grade de
Docteur de l'École Normale Supérieure

Memory for random time patterns in sensory systems

Préparée au
Laboratoire des Systèmes Perceptifs, UMR 8248,
Département d'études cognitives, École Normale Supérieure

soutenue à Paris le 18 Décembre 2017 devant un jury composé de

Daniel Pressnitzer	Directeur de recherche, École Normale Supérieure	Directeur de thèse
Maria Chait	Professeur, University College London	Rapporteur
Laurent Demany	Professeur, Université de Bordeaux	Rapporteur
Laurent Bourdieu	Professeur, École Normale Supérieure	Examineur
Jan Schnupp	Professeur, City University of Hong Kong	Examineur
Simon Thorpe	Professeur, Université de Toulouse	Examineur

Abstract

Time is a universal feature of all information processed by sensory systems, and temporal patterning is often essential for attributing meaning to external stimuli such as speech or music in audition. However, many of the mechanisms needed for temporal processing are still unclear. In this thesis, we investigated one type of mechanism arguably integral to any kind of temporal processing: the formation of novel memories for temporal patterns. We studied mainly the auditory modality, but also other sensory modalities such touch, vision, and electric hearing with a cochlear implant. All experiments used a novel experimental paradigm, adapted from a previous study designed to explore auditory memory of random noise (Agus, Thorpe, & Pressnitzer, 2010). Instead of using noise as the complex stimulus to learn, we used irregular time patterns made of random time intervals delineated by modality-adapted brief energy pulses. In a first series of experiments, we investigated the auditory modality in normal hearing listeners, using click trains as stimuli. We demonstrated, for the first time, a rapid learning of stimuli containing solely temporal cues. In a second series of experiments, we applied the same paradigm to multiple sensory modalities (audition, touch, and vision), using audio clicks, motion pulses to the fingertips, and light to delineate time intervals. We found a qualitatively similar forms of perceptual learning for all three modalities, with rapid learning in all cases, as well as a transfer of learning to touch or vision for patterns learnt initially learnt in audition. In a third series of experiments, we tested hearing impaired listeners stimulated through their cochlear implant with sequences of electrical pulses. We found evidence for preserved plasticity for the rapid learning of time patterns in those listeners. Finally, we present preliminary data using a novel technique for studying the underlying neural mechanisms of rapid perceptual learning. We measured pupil dilation while listeners performed the memory task and observed systematic changes in pupil size with perceptual learning. In conclusion, the thesis shows a remarkable ability of perceptual systems to learn complex time patterns as they re-occur in the environment, and suggests new experimental methods to further study the underlying neural mechanisms.

Resumé

Le temps est une dimension universelle traitée par les systèmes sensoriels, qui est essentielle pour attribuer un sens à des stimuli comme la parole ou la musique pour l'audition. Cependant, les mécanismes requis pour le traitement temporel restent en grande part méconnus. Dans cette thèse, nous avons examiné un type de mécanisme faisant sans doute partie intégrante de tout traitement temporel: la formation de nouvelles traces mnésiques pour l'information temporelle. Nous avons étudié principalement la modalité auditive, mais aussi d'autres modalités sensorielles, comme le toucher, la vision, et la stimulation électrique directe du système auditif périphérique avec un implant cochléaire. Toutes les expériences ont utilisé un nouveau paradigme expérimental, adapté de précédentes études conçues pour étudier la mémoire auditive (Agus, Thorpe, & Pressnitzer, 2010). Au lieu d'utiliser du bruit comme stimulus, nous avons utilisé des séquences d'intervalles de temps irréguliers délimités par de brèves impulsions d'énergie, adaptées à la modalité étudiée. Dans une première série d'expériences, nous avons étudié la modalité auditive chez des auditeurs normo-entendants, en utilisant des trains de clics audio comme stimuli. Nous avons démontré, pour la première fois, un apprentissage rapide de sons contenant uniquement des informations temporelles. Dans une seconde série d'expériences, nous avons appliqué le même paradigme à trois modalités sensorielles (audition, toucher et vision), en utilisant des clics audio, des impulsions de mouvement au bout des doigts, et des flashes de lumière pour délimiter les intervalles de temps dans les différences modalités. Nous avons observé des formes qualitativement similaires d'apprentissage perceptif pour les trois modalités, avec un apprentissage rapide dans tous les cas, ainsi qu'un transfert d'apprentissage au toucher ou à la vision pour des séquences initialement apprises de façon auditive. Dans une troisième série d'expériences, nous avons testé des malentendants stimulés électriquement par leur implant cochléaire avec des séquences d'impulsions irrégulières. Nous avons trouvé des indications d'une plasticité préservée pour l'apprentissage rapide des informations temporelles chez ces auditeurs. Enfin, nous présentons des résultats préliminaires en utilisant une nouvelle technique susceptible de révéler certains des mécanismes neuronaux sous-jacents à l'apprentissage perceptuel rapide. Nous avons mesuré la dilatation pupillaire pendant que les auditeurs effectuaient la tâche de mémoire auditive et observé des changements systématiques de la taille de la pupille avec l'apprentissage. En conclusion, la thèse montre une capacité remarquable des systèmes perceptifs à apprendre des séquences

temporelles complexes lorsqu'elles apparaissent plusieurs fois dans l'environnement, et suggère de nouvelles méthodes expérimentales pour étudier plus avant les mécanismes neuronaux sous-jacents.

Table of Contents

CHAPTER 1 INTRODUCTION	9
1.1 MEMORY FOR AUDITORY INFORMATION.....	9
1.2 TEMPORAL SENSITIVITY	14
1.2.1 Single time interval discrimination and learning in audition.....	14
1.2.2 Single time interval discrimination and learning in touch and vision	17
1.2.3 Single time interval discrimination in electrical hearing	18
1.2.4 Model for single time interval perception.....	19
1.3 TEMPORAL PATTERNS	21
1.3.1 Temporal pattern processing in audition.....	21
1.3.2 Temporal pattern processing in touch and vision	23
1.3.3 Temporal pattern processing in electrical hearing	28
1.3.4 Model for temporal pattern perception	30
1.4 TEMPORAL PROCESSING FOR MULTI-SENSORY INFORMATION.....	32
1.4.1 Modality dependent mechanisms from cross-modal studies	32
1.4.1.1 <i>Audition and Vision</i>	33
1.4.1.2 <i>Audition and Touch</i>	35
1.4.1.3 <i>Further studies across modalities</i>	36
1.4.2 Centralised mechanisms from cross-modal studies	38
1.5 THE AIMS OF THE THESIS	39
CHAPTER 2 AUDITORY MEMORY FOR RANDOM TIME PATTERNS	41
2.1 ABSTRACT.....	41
2.2 INTRODUCTION.....	42
2.3 EXPERIMENT 1: UNIFORMLY DISTRIBUTED TIME PATTERNS.....	46
2.4 EXPERIMENT 2: POISSON-DISTRIBUTED TIME PATTERNS	55
2.5 EXPERIMENT 3: MEMORY TRANSFER AFTER TIME REVERSAL.....	59
2.6 CROSS-CORRELATION MODEL	63
2.7 DISCUSSION.....	67
2.8 CONCLUSIONS	70
CHAPTER 3 MEMORY FOR RANDOM TIME PATTERNS IN AUDITION, TOUCH, AND VISION	77
3.1 ABSTRACT.....	77

3.2	INTRODUCTION.....	79
3.3	EXPERIMENTAL PROCEDURES.....	83
3.3.1	General information common to all experiments	83
3.3.2	Pre-test part I: Setting the masking noise level.....	85
3.3.3	Pre-test part II: Setting the minimum IPI.....	86
3.3.4	Experiment 1	87
3.3.5	Experiment 2.....	88
3.3.6	Statistical analyses	89
3.4	RESULTS	91
3.4.1	Pre-test	91
3.4.2	Experiment 1	93
3.4.3	Experiment 2.....	97
3.5	DISCUSSION	100
CHAPTER 4 MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT		107
4.1	ABSTRACT.....	107
4.2	INTRODUCTION.....	109
4.3	EXPERIMENT	111
4.4	RESULTS	115
4.5	DISCUSSION	118
CHAPTER 5 PUPIL DYNAMICS AS AN INDICATOR FOR AUDITORY MEMORY		121
5.1	ABSTRACT.....	121
5.2	INTRODUCTION.....	123
5.3	EXPERIMENT	125
5.4	RESULTS	129
5.5	DISCUSSION	131
CHAPTER 6 CONCLUSIONS.....		135
6.1	MAIN FINDINGS	135
6.2	PERSPECTIVES	139
REFERENCES.....		142
APPENDIX.....		158

Abbreviations

AFC: Alternative forced-choice

A1: Primary auditory cortex

ANOVA: Analyses of Variances

C: random click trains

CI: Cochlear implant

DL: Difference limen

EEG: Electroencephalography

FixRN: Fixed RN

fMRI: functional magnetic resonance imaging

HL: Hearing level

ICI: Inter-click-interval

IPI: Inter-pulse-interval

ISI: Inter-stimulus-interval

JND: Just Noticeable Differences

LC: Locus coeruleus

LFP: Local field potential

M: Mean

ms: millisecond

MEG: Magnetoencephalography

NH: Normal hearing

P: Random pulse trains

PEG: Ptuallosterior ectosylvian gyrus

RC: Repeated click trains

RefRC: Reference repeated click trains

RefRP: Reference repeated pulse trains

rms: root-mean-square

RP: Repeated pulse trains

s: second

SD: Standard deviation

SOA: Stimulus onset asynchrony

SPL: Sound pressure level

Chapter 1

Introduction

1.1 Memory for auditory information

To make sense of the world, we must be able to recognise and identify patterns as they unfold in our environment. In most cases, we receive a mixture of multiple sensory signals that evolve over time and are concurrently presented. However, we are able to focus on specific sources that contain relevant information. In hearing, for instance, the ability to follow a conversation in a noisy environment is a well-known example, referred to as the "cocktail party" phenomenon (Cherry, 1953). How people segregate or "stream" acoustic signals to acquire necessary information over time has been studied extensively (Moore & Gockel, 2012). The focus so far has been mostly to map how distinct acoustic features of each stream participate in the auditory stream segregation. Effects have been found for frequency range, spatial location, temporal coherences, and timing cues among others (Bregman, 1994; Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2000; Jones, Alford, Bridges, Tremblay, & Macken, 1999; Schnupp, 2006; Shamma, Elhilali, & Micheyl, 2011). Combining information available in these spectro-temporal cues enables listeners to identify each auditory stream and process information effectively. However, one necessary factor that supports this ability but often overlooked is memory (Demany & Semal, 2007). Memory processes have been studied in the scene analysis with what has been termed "schema-based" mechanisms, where listeners have learnt what to listen for in a mixture (Bey & McAdams, 2002). But, if those memory processes were available on a shorter time scale, they could also help to identify re-occurring patterns as they unfold over time (Andreou, Kashino, & Chait, 2011; McDermott, Wroblewski, & Oxenham, 2011; Sohoglu & Chait, 2016).

Recently, a series of studies have reported the possibility of rapid learning for newly presented sounds. In the study of Agus, Thorpe, and Pressnitzer (2010), an

INTRODUCTION

experimental task was introduced where listeners had to report whether stimuli contained a repetitive pattern or not. Patterns were either random Gaussian noise samples, or two abutting identical half-duration noise samples (Figure 1-1).

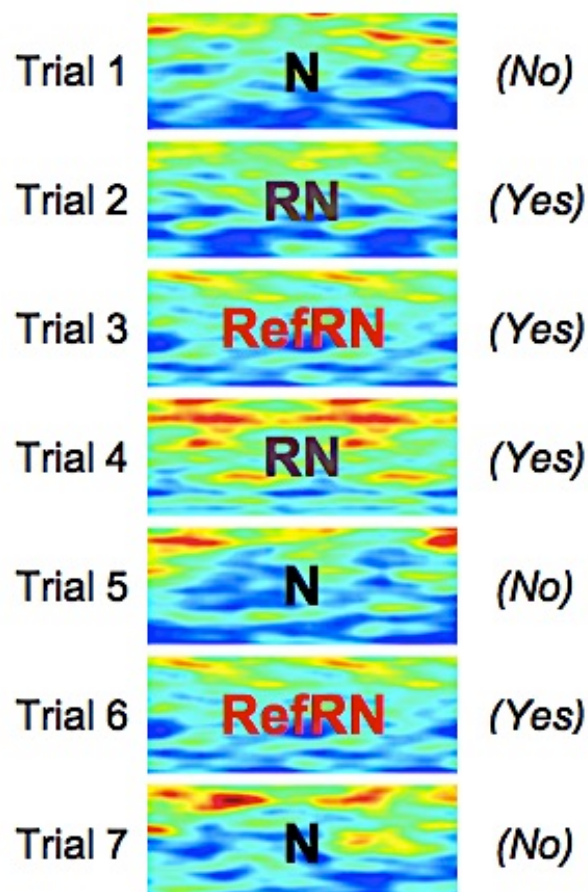


Figure 1-1. Schematic of a repetition detection task. Noise sequences were either random (N) or repeated within a trial (RN). Both N and RN were generated afresh for each trial. One specific RN re-occurred identically across several trials (RefRN). Subjects were only asked to report whether they heard a within-sequence repetition or not. Yes and no responses in a bracket refer to the correct answers for each trial. From Agus et al., (2010).

In more details, in their first experiment, noise samples were either random Gaussian noise for 1 s (simple noise: N) or a 0.5 s random noise token presented twice in a row to create a repetitive pattern within the sequence (repeated noise: RN). Both types of stimuli were generated afresh for each trial and listeners were simply asked to detect a within-sequence repetition, the difference between N and RN. One additional stimulus

type was however also introduced: a specific RN sequence (reference repeated noise: RefRN), which would re-appear identically several times within a test block, unbeknownst to the listeners. The authors measured d' from the signal detection theory (Tanner, Patton, & Atkinson, 1965) as a performance index to compare how well subjects correctly detected the within-trials repetition. If memory of RefRN could be formed from its reoccurrences, subjects should be able to show a better repetition detection performance for RefRN, compared to new RNs.

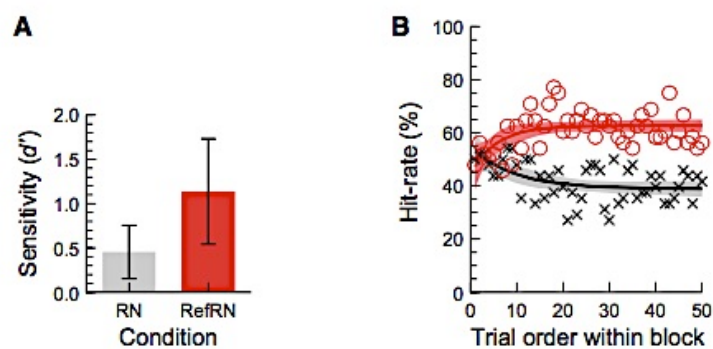


Figure 1-2. Results of the repetition detection task. (A) Higher sensitivity, d' , for RefRN compared to RN indicates that listeners were able to better detect within-trial repetitions for RefRN than RN. (B) Hit-rate of each trial order for RN and RefRN shows that the performance for RefRN increased with its re-occurrences. From Agus et al. (2010).

The results supported this assumption, by showing greater performance for RefRN compared to RN. Importantly, this performance difference did not appear from the first occurrence of RefRN. Rather, performance for both RN and RefRN for their first presentation was about at chance level. The performance for RefRN showed a rapid improvement with its re-occurrences (Figure 1-2). In this experiment, the performance for RN, on the other hand, decreased with number of trials. This was attributed to a criterion effect: more "yes" responses to RefRN, as it became easier to report, induced a bias to provide more "no" responses to RN.

Random noise sequences were chosen in this experiments as a means to investigate primarily auditory perceptual memory, since listeners cannot easily associate particular examples of noise with any extra-auditory information, like semantic labels for instance. These findings thus suggested the existence of fast and robust auditory memory processes for arbitrarily complex auditory signals.

INTRODUCTION

Neural correlates for this form of rapid perceptual learning were subsequently reported by two studies using non-invasive human neuro-imaging tools. The study of Luo, Tian, Song, Zhou, and Poeppel (2013) and the study of Andrillon, Kouider, Agus, and Pressnitzer (2015) used the repetition detection task of noise sequences, following the test paradigm of Agus et al. (2010) as we just described, while recording brain signals using magnetoencephalography (MEG) or electroencephalography (EEG), respectively. The MEG study of Luo et al. did not find any differences in evoked responses between the two stimulus types. Instead, they observed an increase in the phase coherence of responses across trials between 3 – 8 Hz for RefRN when compared to RN. This cross-trial phase coherence found during the presentation of RefRN was more apparent towards later trial orders, suggesting that the difference was due to the learning of RefRN, and possibly mediated by oscillatory brain responses. The EEG study of Andrillon et al. (2015) reported similar findings to those of the MEG study when they presented a "classical" format of noise sequences, observing higher across-trial coherences in 1 – 4 Hz for RefRN, when compared to N. However, they additionally introduced a new format of noise sequences. This new format contained 0.2 s of noise snippets to be repeated 5 times, but, importantly, separated by 0.3 s of random noise which could not be learnt across trials. The motivation for such format was to restrict the time period in the stimulus over which learning and its associated neural correlates could occur. In this condition, they observed evoked potentials, in the form of larger negative responses evoked by RefRN and RN when compared to responses evoked by N. This difference was also observed for the very first snippet of learnt noise in RefRN, implying that the snippet was actually learnt rather than recognised by within-sequence repetitions. The early repeats of the noise snippets in RefRN produced a larger evoked response than in RN, but this difference disappeared at the 5th occurrence of the snippet within RN, suggesting that such a small number of presentations was sufficient to achieve learning. The evoked responses were attributed to plastic changes in secondary areas. Another brain imaging study, this time using functional magnetic resonance imaging (fMRI), suggested that such changes involved secondary auditory region (Planum temporale) as well as hippocampus (Kumar et al., 2014). In summary, these results show that the rapid perceptual learning observed during behaviour has rapid correlates in sensory regions (probably secondary auditory cortex), with a contribution of hippocampus.

Behavioural and neural data thus revealed an impressive characteristic of auditory memory: the ability to learn arbitrarily complex sounds, rapidly and in a long-lasting way

(Agus et al. 2010; Viswanathan, 2016), which would be highly desirable features to learn everyday sounds. However, it is unclear what feature(s) of the auditory stimuli contributed to this type of memory. From additional experiments, it was suggested that spectral information was the main cue used by listeners to achieve perceptual learning (Agus et al., 2010; Agus, Jeannou, & Pressnitzer, 2013; Viswanathan, 2016; Viswanathan, Rémy, Bacon-Macé, & Thorpe, 2016). For example, Agus et al. (2010) presented time-reversed versions of the RefRN stimuli to listeners, after they had learnt the original RefRN. They found that listeners could perform the repetition-detection task as well on these transformed versions as on the original RefRN, suggesting that the cues to learning were robust to time reversal. Spectral cues, such as the brief occurrence of a random salient frequency or spectral pattern, would be resistant to time-reversal. Furthermore, Agus et al. (2013) asked musicians to annotate RefRN into musical notation, and musicians spontaneously transcribed the stimuli to short melodies, in a consistent manner for each musician and noise stimulus. Finally, Viswanathan, Rémy, Bacon-Macé, and Thorpe (2016) divided RefRN snippets after learning into smaller segments and shuffled the order of these segments. Interestingly, listeners still showed robust performance to detect repetitions on the shuffled RefRN, suggesting again the use of brief spectral cues. In summary, the current evidence strongly suggest that listeners isolated brief random spectral cues to achieve rapid perceptual learning of random noise stimuli.

This led to the first main question for the thesis: can auditory *temporal* information, instead of spectral, be learnt by similarly rapid memory processes? Temporal information is central to the auditory modality, as sound waves evolve by necessity over time. The auditory modality is also thought to hold the greatest temporal sensitivity among the senses (e.g., Goodfellow, 1934). It would thus be important to know whether the rapid memory processes observed for noise so far would generalise to other types of sounds, in particular sounds containing the important class of temporal auditory cues. In addition, it can also be remarked that temporal information is ubiquitous in sensory systems. Sensory information is always presented over time, with information often available in its temporal patterning, even for touch such as feeling a texture or for vision such as observing a dynamic visual scene. This opens a second question: would rapid learning of temporal information be unique to audition, or also observed for different sensory modalities?

The present thesis aimed to investigate the memory of purely timing information

INTRODUCTION

mainly in the auditory modality but also in other sensory modalities such as touch, vision, and electrical hearing with a cochlear implant. In the remaining parts of the Introduction, we will briefly survey the literature on temporal processing in audition and in the other modalities tested in the experiments. Note that two main experimental chapters are organised in journal publication format, so they all also include a brief literature review. In the first chapter, we will focus on studies not discussed in details in the experimental chapters.

1.2 Temporal sensitivity

Basic understanding of temporal processing starts from how sensitive we are on simple time judgment tasks. Estimating the duration of a sensory input or an interval marked by two sensory inputs is the simplest form of time judgment that has been extensively studied in the literature. Temporal sensitivity is then often described by the minimum threshold value that enables to correctly discriminate two stimuli of different duration. In this section, we first address temporal sensitivity across sensory modalities, starting from audition (1.2.1), then touch and vision (1.2.2), and a relevant model to explain the time interval judgments (1.2.3).

1.2.1 Single time interval discrimination and learning in audition

The first question regarding auditory time perception concerns the minimum duration change required for detecting the difference between two durations. Such a difference limen (DL), which can also be referred as a just noticeable difference (JND), is a key notion to quantify temporal acuity. The DL is generally measured by using a comparison task between two different intervals, with either a fixed or roving standard time interval that has been compared to a second interval with different duration. The intervals can be of two types: either filled intervals or empty intervals (Figure 1-3).

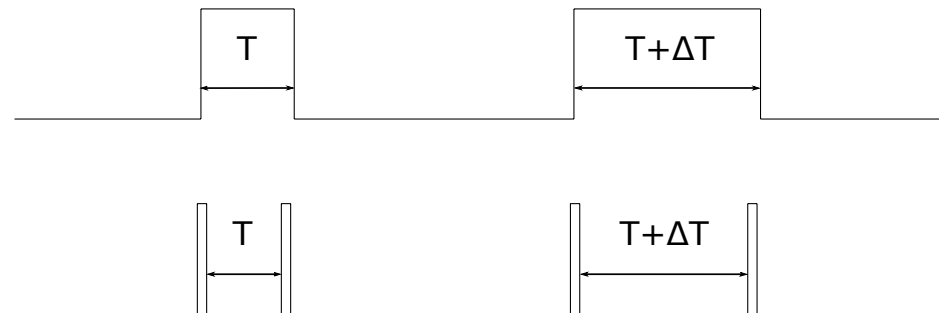


Figure 1-3. Example stimuli used in time interval discrimination tasks. The temporal information to compare is either the duration of sustained sensory stimuli (filled intervals, top), or the duration of silent gaps marked by two brief stimuli (empty intervals, bottom).

Filled intervals are continuous signals presented for certain duration. Empty intervals are marked by brief sensory onset and offset signals determining the duration of silence between them. Usually, a pair of either filled or empty intervals with different durations is presented, which can vary based on a number of features.

One of the early studies, for example, designed a duration discrimination task by presenting one standard filled interval (T) and another filled interval of either the same length (T) or the length increased by an additional quantity ΔT ($T + \Delta T$) (Creelman, 1962). The study reported that listeners could accurately detect duration differences between the two filled intervals over the range tested consisting of base durations from 20 ms to 320 ms, with a performance decrease for longer T s when ΔT s was fixed at 10 ms. The study also reported that the performance improved with bigger ΔT s for the fixed T .

Abel (1972) used a pair of empty intervals marked by two noise bursts, to measure JNDs on a range of base durations. A standard interval T that varied from 0.63 ms to 640 ms was compared with another interval increased by ΔT and the JND was determined per each standard time gap. Abel found that listeners showed better discrimination abilities when markers were shorter and had higher amplitude values. The ratio of $\Delta T/T$ remained similar across standard intervals, of about 10%, with an exception of T below 5 ms where performance was worse with a greater variability. The finding of a non-monotonic Weber fraction across different T could imply different strategies for different based durations. Abel concluded that the Weber fraction was constant at least for the interval duration between 5 and 640 ms.

INTRODUCTION

The perceptual duration of empty intervals and filled intervals has also been found to differ. Some studies observed that the perceived duration of filled intervals was longer than the perceived duration of empty intervals, for equal physical duration (Allan, 1979; Goldstone & Goldfarb, 1963). Furthermore, temporal judgments on filled intervals tend to be more accurate than on empty intervals (Abel, 1972; Craig, 1973; Rammsayer & Lima, 1991). Perception of timing information could also differ based on the task, e.g., between a reminder task (a standard interval is always presented first then followed by a comparison interval) compared to a 2-alternative-forced-choice (2AFC) task with roving (Lapid, Ulrich, & Rammsayer, 2008; Rammsayer, 2014). Therefore, there is no single estimate of temporal acuity even within a modality.

Another type of task involving time interval discrimination is perceptual learning. Several experimental have trained listeners to discriminate a certain base duration interval, and then tested for the generalisation of temporal learning across different stimulus parameters. Wright, Buonomano, Mahncke, and Merzenich (1997) found that the learning was specific to a given time interval, but generalised to changes on spectral cues used to delineate the time interval. In their study, listeners were asked to do the interval discrimination task between a standard time interval and comparison intervals with varying duration. Listeners were trained on one specific standard interval presented with a fixed frequency tone. After a training session, listeners showed similar performance in the discrimination task with the same standard interval but with tones at a different frequency. However, they did not show as good performance on other time intervals. This indicated that the learning of time information was specific to the absolute value but could be generalised for different spectral cues. Karmarkar and Buonomano (2003) also found a generalisation of learned timing information over frequency cues. In their study, listeners received training on the interval discrimination for short intervals with lower frequency and long intervals with higher frequency, which would have associated each interval with a certain frequency. Regardless, listeners showed a performance improvement when the association was reversed, i.e. for short intervals with high frequency and long intervals with low frequency. These results further suggest that spectral and temporal information processing are independent from each other as far as perceptual learning is concerned, which is relevant to our question of learning with spectral versus temporal cues.

1.2.2 Single time interval discrimination and learning in touch and vision

In other sensory modalities, such as touch and vision, fewer studies have been conducted regarding time interval perception. Studies in each modality have different focuses. In touch, studies largely focus on a perceptual variability based on the "shape" of stimuli. For example, temporal gap detection in either sinusoids or noise bursts has been found to differ. Van Doren, Gescheider, and Verrillo (1990) reported that gap detection thresholds were higher when tactile stimuli were noises compared to sinusoids, especially when the gaps was short, below 100 ms.

In vision, one of the early studies reported that the minimum interval duration required for a clear perception of the gap between two visual stimuli at a behavioural level could go up to 80 ms (Nilsson, 1979). As we have seen in audition, perceptual differences between durations marked by filled and empty intervals are also found in vision. Based on Grondin (1993), discrimination of time duration for empty intervals was superior to filled intervals, regardless of standard duration length, either 50 ms or 250 ms. Temporal perception in different visual fields further show different abilities to perceive timing intervals. Temporal processing in the periphery, for instance, is known to be faster than at the fovea, (i.e., better at detecting smaller gaps between two flashes: Allen, Hees, & Nordby, 1998; Seiple, Greenstein, Holopigian, & Carr, 1988). This difference between areas of the visual field could be due to the larger size of peripheral photoreceptors compared to foveal receptors (Tyler, 1985). In both touch and vision, the temporal sensitivity seems to largely vary based on features of stimuli and test settings.

The difference of the temporal sensitivity across modalities can be compared when the same kind of discrimination tasks was used. Generally, the auditory modality has been found to be the most sensitive for temporal gap discrimination, as well as reproduction of timing information, when compared to vision and touch (Gault & Goodfellow, 1938; Goodfellow, 1934; Lechelt, 1975). One of the earliest studies by Goodfellow (1934) compared humans' abilities to discriminate the lengths of temporal gaps delineated by auditory clicks, visual light flashes, or tactile stimulation. Each trial contained 3 brief sensory stimuli in one of the modalities, with one standard interval of 1 s and the other interval varying from 1 s to 1.3 s. Three tasks were tested, with subjects having to either respond whether the second interval was longer than first one, respond whether the lengths of the two intervals was different, or reproduce intervals. In all cases,

INTRODUCTION

the differential thresholds were the lowest for audition (average 70 ms), the highest for vision (average 115 ms), and intermediary for touch (average 95 ms). Overall, the quantitative difference on the sensitivity seems to exist and this difference raises a question on the mechanisms of temporal processing, whether they are centralised or modality-dependent. Studies regarding this issue will be further reviewed in section 1.4.

Finally, there are also experiments involving perceptual learning of a single interval in modalities other than audition, with similar questions of transfer of learning to various stimulus parameters (Nagarajan, Blake, Wright, Byl, & Merzenich, 1998; Sathian & Zangaladze, 1997; Spengler et al., 1997). In the study of Nagarajan et al. (1998), an interval discrimination task was used by presenting two pairs of tactile pulses, one with a fixed duration (referred as the "base interval") and the other with a duration that was longer than the base interval (the "target interval"). Tactile pulses were presented either to the palm of the hand or the fingertip. Participants performed a pre-test session, followed by 10 – 19 days of training, and a post-session. Thresholds for discriminating the base interval and the target interval decreased along the training, showing perceptual learning of the time interval presented in the tactile modality. The learning generalised to the contralateral skin location in subjects who received pulses on the palm and to adjacent fingers for subjects who were trained on the finger, also showing some independence of the learning over stimulus parameters.

1.2.3 Single time interval discrimination in electrical hearing

Cochlear implant (CI) is the only successful sensory prosthesis that connects external sensory inputs to nerve fibres to rehabilitate the sensory process at present, for people with sensorineural hearing loss to restore their hearing abilities. The placement of electrodes from apex to basal area of the basilar membrane in the cochlea covers the frequency range and enables place coding of frequency information. Through electrodes, electrical signals stimulating nerve fibres are delivered, forming the electrical hearing mechanisms. Due to a limited number of electrodes, CI listeners' hearing ability for pitch perception is poor compared to NH listeners. To subsidise limitations in the spectral domain due to a relatively small number of electrodes, this electrical hearing system heavily rely on timing information of electrical signals. Unlike a degraded ability for frequency-related hearing activities for CI listeners (e.g., Kong, Cruz, Jones, & Zeng,

2004) studies have found comparable performance for time interval perception between CI and NH listeners. For example, Shannon (1989) used a 2AFC gap detection task, marked by either sinusoidal or biphasic pulsatile electrical waveforms, to investigate the thresholds of CI listeners. In general, Shannon found the gap detection thresholds as 20 – 50 ms for markers with low-level (quiet) and 1 – 5 ms for markers with high-level (loud), which were similar to the thresholds found in NH listeners (e.g., Penner, 1977). Thus, the temporal sensitivity between NH and CI listeners seems to be similar. Studies regarding temporal processing that could aid listening behaviours of CI listeners when receiving a series of pulses will further be addressed in section 1.3.3.

1.2.4 Model for single time interval perception

The scalar expectancy theory (SET) is the most common model explaining the processes of time interval perception. It is described by a clock process (Figure 1-4), more specifically, a pacemaker generating clock ticks and a switch emitting the clock ticks toward an accumulator in memory (Gibbon, 1977; Gibbon, Church, & Meck, 1984; Rammsayer & Ulrich, 2001; Treisman, 1963).

INTRODUCTION

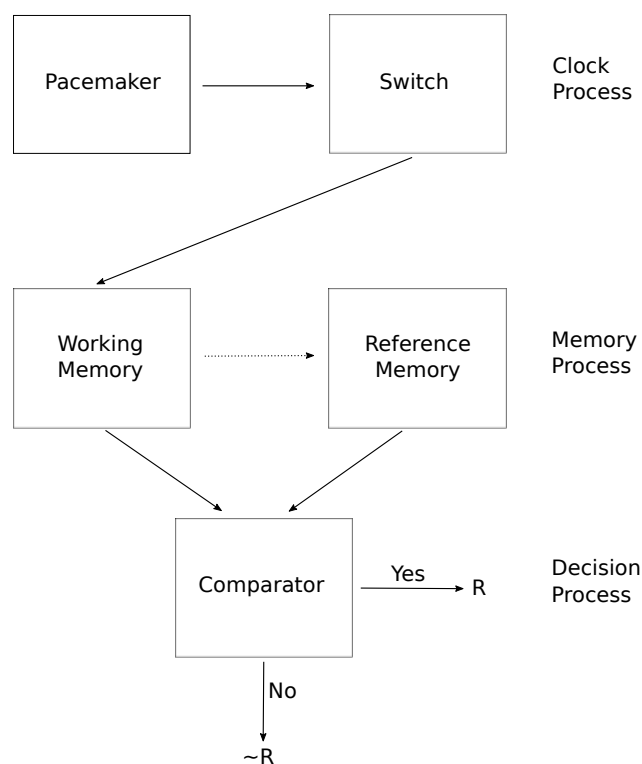


Figure 1-4. Schematic to explain the SET process. From Gibbon, Church, & Meck (1984).

Gibbon et al. (1984) investigated this model to explain the variance of timing judgments. They reported that the variability of temporal estimates is generated by different sources, which are defined as scalar. This scalar variability would generally follow the Weber law, proportional to the mean subjective time. Then, a distribution of the timing judgment would remain similar for a wide range of time durations. This theory can be applied to a temporal generalisation (Church & Gibbon, 1982): After a training with a reinforcement for a given time duration T , animals' responses would peak at T and gradually decrease for both shorter and longer duration to T . The distribution of responses, which can be referred as the scalar variability, would be similar across different T s, confirmed by its superposition when computed in relative time.

This theory, originally developed for understanding animal data, can be applied to investigations of human data (Allan, 1998; Malapani & Fairhurst, 2002; Penney, Gibbon, & Meck, 2000; Wearden, 2003). For example, Wearden (1992) conducted a temporal generalisation study in humans. After exposure to a "standard" tone sequence with each tone of fixed duration, human listeners were asked to respond whether individual target tones lasted for the same duration as the standard or not. Standard durations varied from

400 ms to 700 ms. Also, target durations were linearly or logarithmically scaled with the corresponding standard durations. The authors found an asymmetry of "yes" responses, i.e. there were more positive responses for target that were longer than the standard duration compared to targets that were shorter, and they found decreasing performance for longer standard durations. Nevertheless, overall performance across different standard durations remained qualitatively similar. Performance showed a peak of "yes" responses for the standard duration and decreased for longer and shorter durations, which shaped a similar distribution having the same relative ratio based to the standard duration. The findings confirmed the time judgements followed the Weber law. Similar variability for time judgment performance across different standard stimulus durations could further imply that the similar scalar property was applied. The SET theory has thus been useful for understanding how single time interval information is treated and processed.

1.3 Temporal patterns

Studies introduced in the previous section dealt with understanding the simplest type of information in the time domain, a single time interval. However, in the general case we rather encounter sensory information containing a series of time intervals, and the relevant information is in the temporal patterning as well as in the estimation of single durations. In this section, we will review studies of temporal pattern processing in audition (1.3.1), in touch and vision (1.3.2), and a relevant model to explain pattern processing (1.3.3).

1.3.1 Temporal pattern processing in audition

Music and speech, two of the most common types of natural auditory stimuli, contain meaningful information in the temporal patterning of sounds, such as the timing of each phoneme or note, and their respective duration. Various aspects of temporal information processing have been studied for sound sequences.

How humans treat intervals of different lengths or deviants within a series of temporal patterns has been one focus of previous temporal processing studies. Performance in the discrimination task can differ based on the duration of intervals in the

INTRODUCTION

sequence or even where a time deviant occurs. Hirsh et al. (1990) introduced a time pattern comparison task where participants were instructed to indicate which one of two comparison tone sequences contained a deviant interval at a certain position, compared to a regular standard tone sequence (Figure 1-5).

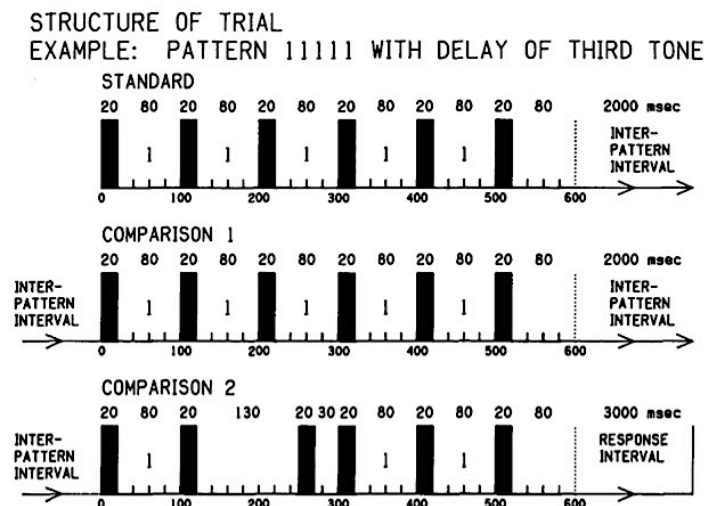


Figure 1-5. Schematic of a tone sequence comparison task. After listening to the standard tone sequence (top), listeners were asked to respond which of a first or second comparison tone sequence contained a deviant interval (middle or bottom). The position of the deviant within the sequence was indicated to listeners before each run. From Hirsh et al. (1990).

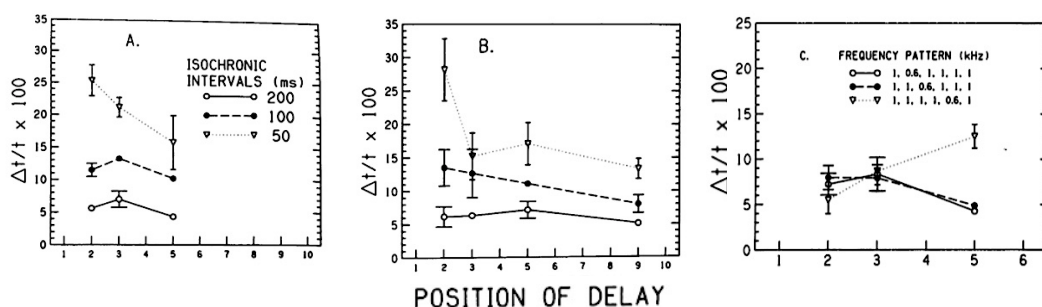


Figure 1-6. Relative DLs in the sequence comparison task. Temporal sequences contained either 6 tones, with 5 intervals (left, A) or 10 tones with 9 intervals (middle, B). For sequences with longer intervals (100, 200 ms), no reliable effect of deviant position on DL was found. For the fast sequence with shorter intervals (50 ms), an early position of the deviant worsened the performance. In the other experiment, temporal sequences had a tone frequency change at the position of the deviant (right, C). This frequency change worsened the performance only when the deviant was presented at a late position in the sequence. From Hirsh et al. (1990).

The main results were that the sequence structure affected the time judgment task of the single deviant interval. More specifically, the task performance remained similar regardless of the deviant position for tone sequences with 100 or 200 ms of inter-onset intervals, whereas the performance was poorer when the deviant was presented at an early position within the sequence with 50 ms intervals (Figure 1-6, left and middle). Temporal sequences with longer intervals produced relatively consistent temporal processing performance compared to faster sequences. Furthermore, when there was a pitch change of the tone creating the deviant interval at a later position within the patterns, the performance was disrupted, implying that temporal processing could be affected by spectral cues in the context of temporal patterns (Figure 1-6, right).

More structured temporal patterns can create rhythmic information, that is, subjected accents on parts of the sequences that depend among other things on the predictability of the sequence (Povel & Essens, 1985). Rhythm has been shown to be a crucial factor for efficient discrimination of timing patterns compared to irregular timing patterns (Bouwer, Van Zuijen, & Honing, 2014; Large & Jones, 1999). For instance, Drake and Botte (1993) reported that JNDs were lower for sequences with regular intervals compared to irregular intervals. They also found that musicians were better at detecting tempo change, and musicians also showed lower JNDs for interval discriminations compared to nonmusicians. This may suggest an effect of familiarity on temporal pattern discrimination, with distinct neural mechanisms for familiar rhythmic patterns. For instance, according to the study of Sakai et al. (1999), temporal patterns regarded as "metrical" rhythms activated left premotor, parietal areas, and the right cerebellar anterior lobe. Patterns containing "non-metrical" intervals created with non-integer values of the interval ratio induced brain activations in right prefrontal, premotor, parietal areas, and bilateral cerebellar posterior lobe.

1.3.2 Temporal pattern processing in touch and vision

Time pattern processing in touch in human has been quantified mostly by measuring the ability to discriminate different rates of tactile patterns. In such studies, it has been shown that human tactile perception can distinguish tactile rates in the range of 2 pulse-per-second (pps) to 300 pps (Sherrick, 1985). Stimulus parameter may influence the estimate

INTRODUCTION

of rate discrimination acuity. Above a rate of 40 Hz, the rate detection threshold decreased when a size of a contactor area increased (Verrillo, 1965; 1985).

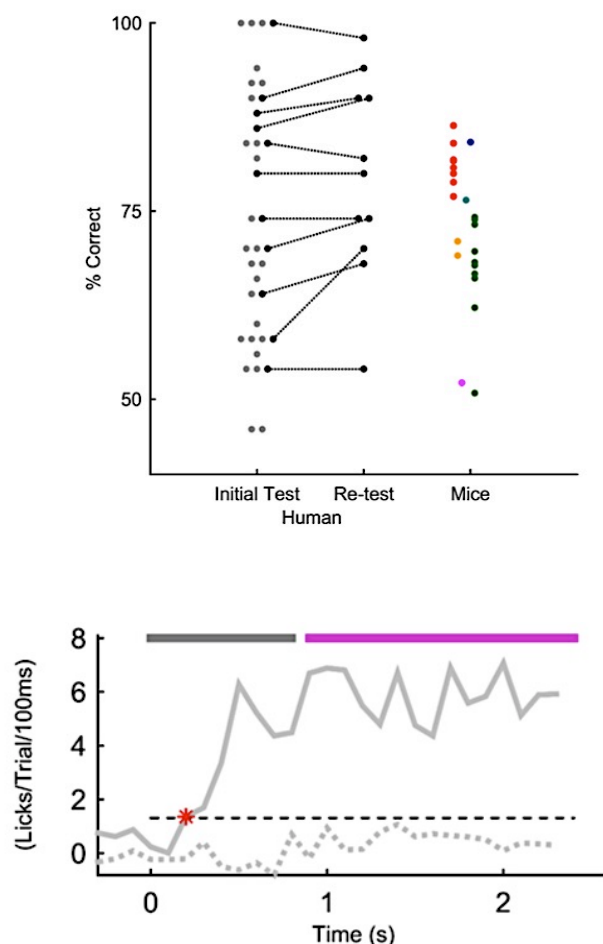


Figure 1-7. Results of a tactile pattern discrimination task. (Top) After a set of training, both humans and mice were able to correctly identify the target pattern with similar performance levels. (Bottom) Plots of the difference in lick rate between trials with the target pattern and trials with non-target patterns (sequence with target pattern segments scrambled in the hundreds of milliseconds range, solid grey line). The dashed grey line represents differences in average lick rates from 50 randomly drawn trials and the dashed black horizontal line is the 95% confidence interval of lick rate values from the 50 repeats. An asterisk indicates where the meaningful change occurs between the actual difference in the lick rate and the difference of randomly drawn trials. The top thick grey horizontal bar indicates the sequence presentation period and the top thick fuchsia bar indicates the response period when the water was actually available. From Bale et al. (2017).

Perceptual learning of sequences has also recently been investigated with tactile stimuli, in a setting inspired by the study of Agus et al. (2010). Bale et al. (2017)

presented tactile patterns consisting of different amplitude modulations, regarded as "syllables", on whiskers for mice and fingertips for humans. In all cases, participants were trained to distinguish a reference sequence from other sequences that were created by scrambling the amplitude syllables. The result showed that both mice and humans could learn to discriminate reference temporal patterns from other patterns with shuffled syllables (Figure 1-7, top). Responses of mice were made as soon as changes in a sequential order were detected, even before the response period (Figure 1-7, bottom), suggesting an immediate detection of the reference sequence for its presentation and an efficient form of learning.

Furthermore, studies regarding a transfer of learning have been performed with temporal sequences in addition to the single time interval studies reviewed in previous sections. Harris, Harris, and Diamond (2001) presented a set of stimuli differing in either vibration speed (fast or slow), punctuate pressure (hard or soft), or roughness (rough or smooth).

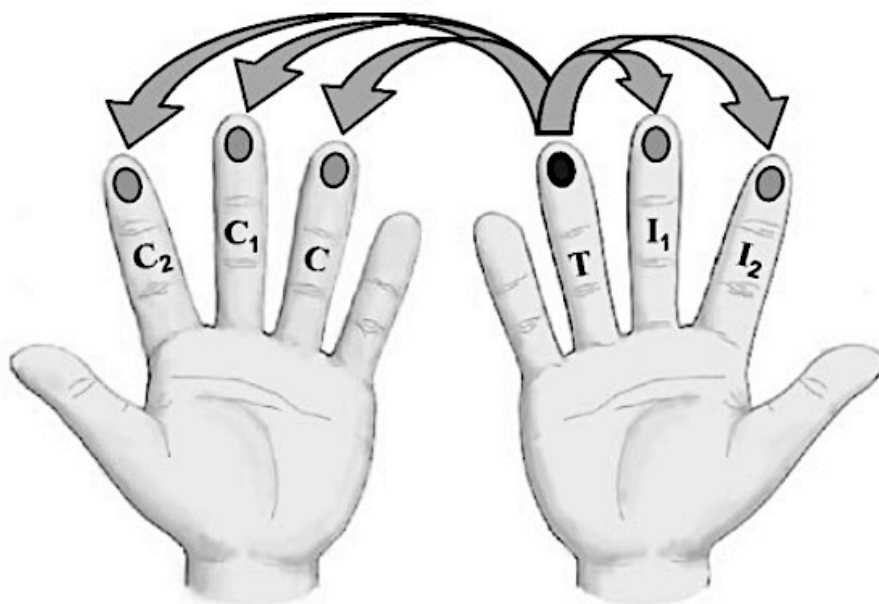


Figure 1-8. An example of tested fingers in the tactile discrimination task. After training on one finger (T), neighbouring fingers in an ipsilateral position as well as parallel fingers in a contralateral location were also tested on the same discrimination task to investigate generalisation of tactile learning. From Harris et al. (2001).

INTRODUCTION

Participants were trained on a designated finger that was different for each task, and were tested after the training on different fingers: the trained finger, adjacent fingers, and fingers on the opposite hand (Figure 1-8). The transfer was stronger for the first next finger as well as for the topographically matched finger on the opposite hand, for the roughness and pressure tasks. Harris and colleagues suggest that the strong transfer would only be found when the trained area and the test area were topographically matched. They did not observe a transfer of vibration speed learning but the test frequency (between 9 Hz and 10 Hz) differed from the training frequency (between 8 Hz and 10 Hz), which could have limited the effect inducing the transfer. As we have seen in single time interval transfer studies in audition as well as touch, these findings suggest that the transfer is not limited to a single interval, but also can be found for a more complex form of time intervals: multiple time intervals presented as a group.

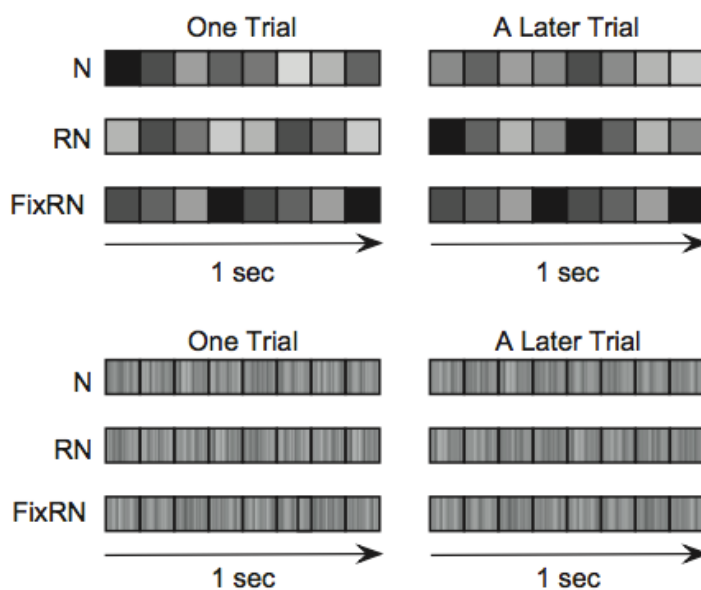


Figure 1-9. Examples of 1D (top) and 2D (bottom) visual contrast noise sequences, either random for all eight squares (N) or first four random squares repeatedly presented (RN). Unlike N and RN that change contrast levels of square for each trial, FixRN, which has the same format as RN, re-occurs for several trials with identical features of its squares. From Gold et al., (2014).

Studies of temporal pattern processing in vision are less common and time-related processing in the visual domain is generally known to be poor. The study of Gold, Aizenman, Bond, and Sekuler (2014) explored a memory formation of random spatio-temporal visual sequences, by adapting the paradigm of Agus et al. (2010). Gold et al.

produced contrast noises as visual stimuli. Each sequence consisted of eight 128 x 128 pixel squares either with eight different contrast levels for each square (1D), or different contrast samples to each pixel within each square that are again different for each square (2D). Each square was presented for about 133 ms, followed by subsequent square presentations in a time series, for each visual sequence to last around 1 s. The same within-sequence repetition detection task from the study of Agus et al. (2010) was used, by presenting random noise sequence (N), repeated noise sequence (RN), and one specific RN that re-occurs for several trials with the same features (FixRN, Figure 1-9). They observed a gradual performance improvement for FixRN, the sequence that re-occurred for several trials, compared to relatively constant performance for RN, although the gain of learning was smaller and less rapid, when compared to the result of auditory noise learning that Agus et al. observed (Figure 1-10).

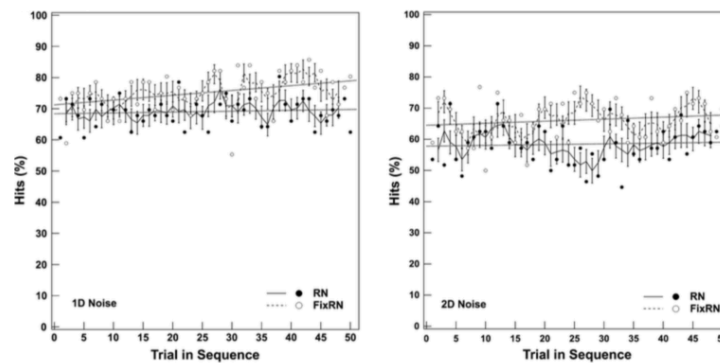


Figure 1-10. Performance of each trial for RN (filled circles and solid line) and FixRN (empty circles and dashed lines), for 1D noise (left) and 2D noise (right). Error bars indicate $\pm 1 SD$. From Gold et al. (2014).

Furthermore, Gold et al. introduced a new type of FixRN, by presenting the second half of FixRN in a reversed order. They did not observe any learning effect for the reversed FixRN. These findings suggest that incidental learning of random noise sequences was poor in vision compared to that in audition, and was susceptible to temporal information.

Barakat, Seitz, and Shams (2015) also studied perceptual learning of visual temporal patterns. They introduced a set of two series of visual stimuli, and asked participants to respond whether two stimuli were the same or different. The task

INTRODUCTION

contained a pre-test training, post-test 1, another set of training, and a final post-test to study if participants could improve their task performance throughout the training sessions. Participants who received the training session with visual stimuli did not show any performance improvement. Interestingly, on the other hand, participants who received the training session with auditory stimuli or audio-visual stimuli (temporal patterns presented in both auditory and visual modalities) showed significant performance improvements, which is an indication of perceptual learning. Although only a few, the studies in vision so far suggest that temporal processing in vision is relatively poor, especially when compared to audition. However, it is unclear whether this difference indicates that there are distinct temporal processing mechanisms for individual modality or not. More detailed review on different performance across modalities on tasks regarding will be introduced in section 1.4.

1.3.3 Temporal pattern processing in electrical hearing

As we have reviewed earlier in section 1.2.3, the sensitivity for timing information of CI listeners is comparable to that of NH listeners. In fact, CI listeners actively rely on temporal processing to acquire pitch information, referred as rate coding. The ability of temporal processing for CI listeners has been studied by using a rate discrimination task. For example, Zeng (2002) delivered pulse trains of different frequencies ranging from 50 Hz to 500 Hz to either the most apical or the most basal electrode. CI listeners received a 3-AFC task for rate discrimination as well as an absolute magnitude estimation to match given pulse rates to the numerical pitch value. The average frequency DL for the standard frequency of 50 Hz was 12 Hz, and the DL gradually increased with the increase of the standard frequency (Figure 1-11, left). Also, an increase of pitch estimation along with faster rates was found up to 300 Hz then saturated for rates above 300 Hz (Figure 1-11, right). The performance did not highly depend on which electrode the pulses were presented. The result indicated the independence of rate coding from the place coding, and a perceptual limitation of pitch discrimination from the rate coding at 300 Hz.

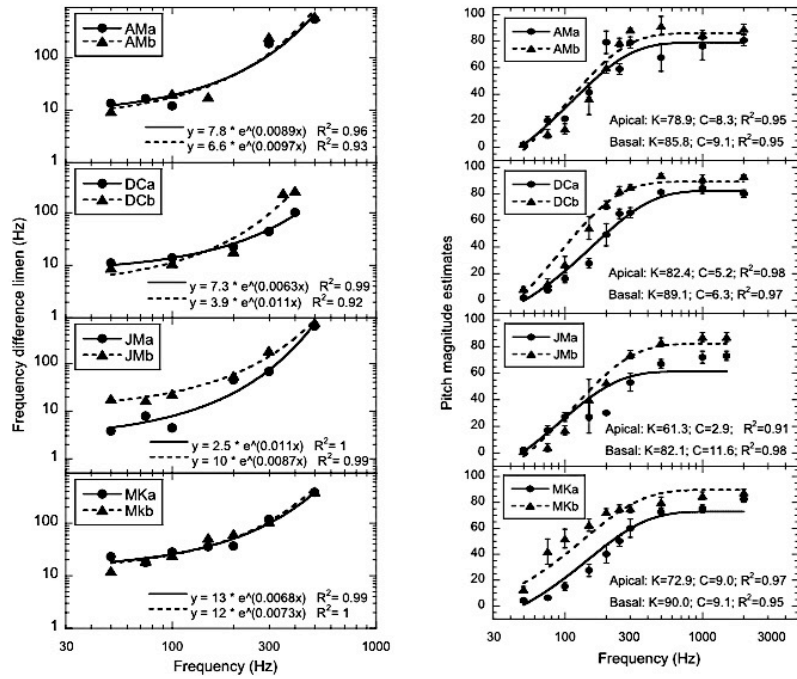


Figure 1-11. Results of a rate discrimination task (left) and a pitch estimation (right) of four CI listeners, from 50 Hz and 500 Hz presented at either the most apical electrode (circles) or the most basal electrode (triangles). From Zeng (2002).

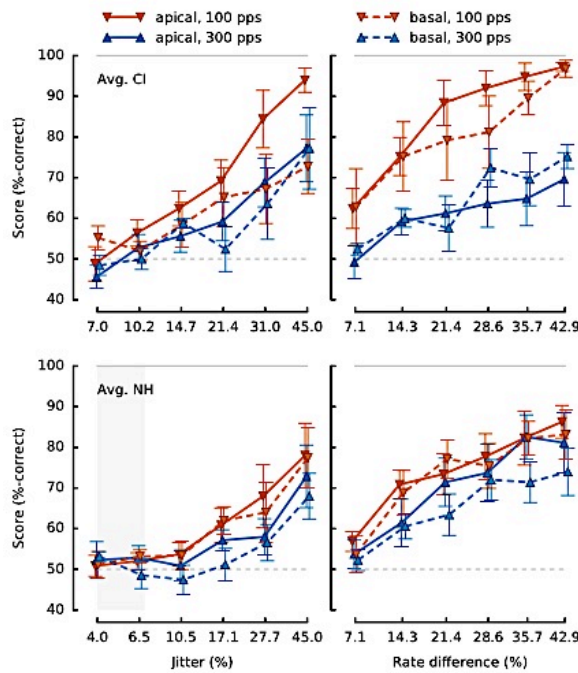


Figure 1-12. Average performance for a jitter detection task (left) and a rate discrimination task (right), at the rate of 100 Hz (red) and the rate of 300 Hz (blue), for CI listeners (top) and NH listeners (bottom). Solid lines indicate pulses sent to the apical electrode for CI or "high" frequency region (3900 – 5400 Hz) for NH listeners, and dashes lines indicate pulses sent to the basal electrode for CI or "very high" frequency region (7800 – 10,800 Hz). From Gaudrain et al. (2017).

INTRODUCTION

Gaudrain, Deeks, and Carylton (2017) used a jitter detection task in addition to the rate discrimination task, to study if the limitation of pitch perception for CI listeners was due to inaccurate coding of time intervals within pulse trains through the electrical hearing mechanisms. They used three-interval, 2AFC task, by presenting a standard stimulus first, followed by another standard stimulus and test stimulus in a random order, and asking participants to choose one that is different from the standard stimulus between the second and the third stimuli, with 100 Hz and 300 Hz as standard frequencies. They then compared the task performance of CI listeners to that of NH listeners. Although the rate discrimination performance of CI listeners had some differences between 100 Hz and 300 Hz, overall results showed that the performance of CI listeners and NH listeners was highly comparable, for both the jitter detection and the rate discrimination tasks (Figure 1-12). Overall, regardless of perceptual limitation on temporal pitch perception in certain circumstances, the findings suggest that the electrical hearing mechanisms persevere efficient temporal processing.

1.3.4 Model for temporal pattern perception

Whilst SET, introduced in section 1.2.3, focused on a single time interval, the dynamic attending theory (DAT) has been developed for temporal sequences, hypothesising an attentional entrainment to a temporal structure. Suggested by Jones and Boltz (1989), DAT emphasises the role of expectancy, expressed by "self-sustaining" oscillators. According to the theory, external rhythms induce attentional rhythms that follow the regularities of timing events. Once the regularity is entrained, the expectancy of forthcoming events can be developed. Internal oscillators are created by the entrainment and adaptation to the dominant period of temporal information in the environment, manipulated in experiments through what is called the "adaptation phase". The oscillators adjust their period to the external stimulus rhythm.

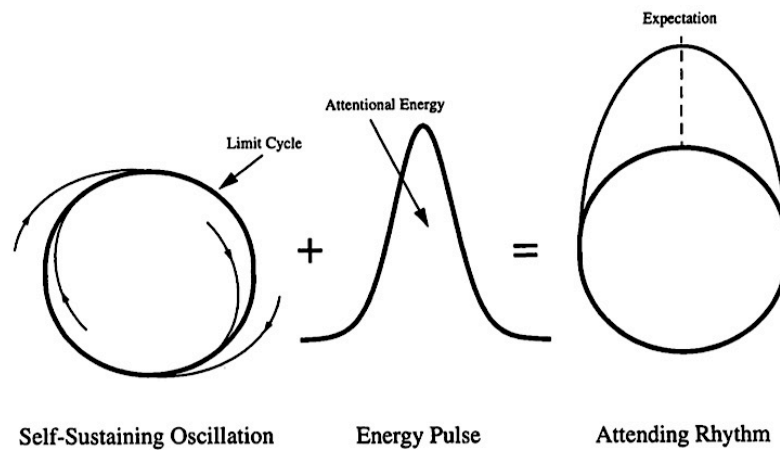


Figure 1-13. Schematic of the attending dynamic model. A self-sustained oscillator and attentional energy pulse compose an attending rhythm. From Large and Jones (1999).

Experimental findings and model predictions further developed by Large and Jones (1999) tested this theory (Figure 1-13). They found better discrimination of time intervals when targets were preceded by the presentation of a stationary rate pattern. Their model predicted this finding by the coupling of attentional and external rhythms.

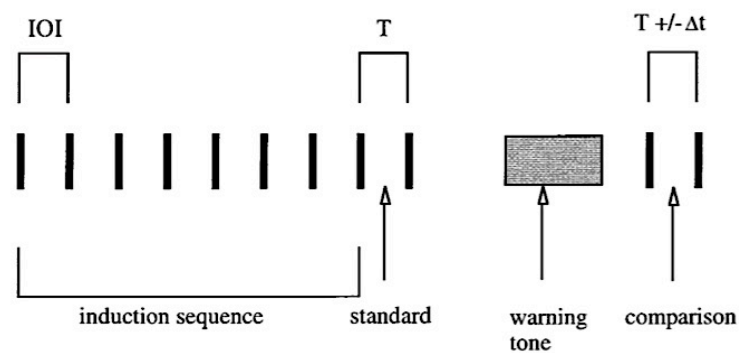


Figure 1-14. Illustration of the timing interval discrimination task. From Barnes and Jones (2000).

Further predictions are that more exposures to standard time information would improve judgments of target timing (McAuley & Kidd, 1998; Vos, van Assen, & Frañek, 1997) and discrepancies between context time events and attending time would reduce the performance (Barnes & Jones, 2000). The study of Barnes and Jones presented

INTRODUCTION

context sequences with 600 ms of inter-onset-intervals (IOIs) followed by a varying standard interval to be compared with a target interval (Figure 1-14). They found that standard intervals that were deviant from the context sequence were detrimental to the interval judgment, compared to when the standard interval remained the same as the IOI of the context sequence.

DAT requires preceding contexts to entrain one's attention. This theory is thus directly applicable for a temporal information that introduces "rhythmic" patterns, which may not necessarily be the general case for environmental stimuli. There are nevertheless studies showing neural evidence in support of DAT, in particular neural entrainment generated by a series of events with rhythms and conventional musical metric frequencies that could support the role of timing inputs for time perception (Besle et al., 2011; Nozaradan, 2014).

1.4 Temporal processing for multi-sensory information

So far, we have reviewed the temporal sensitivity data from single interval comparison tasks as well as some studies investigating a series of time intervals, but always in an individual modality. However, daily scenes often consist of concurrent multi-sensory events, rather than one sensory input at a time. Thus, how we process combined information efficiently is another key issue in timing research. A related debate is whether temporal processing is due to centralised or distributed mechanisms (e.g., Ivry & Schlerf, 2008; van Wassenhove, 2009; Grondin, 2014). A centralised processing could be applicable to all modalities and induce strong interactions, whereas differences across modalities would be predicted from distributed processing. In this section, we will review studies of multi-sensory temporal processing that suggest either modality-dependent mechanisms (1.4.1) or general mechanisms across modalities (1.4.2).

1.4.1 Modality dependent mechanisms from cross-modal studies

We will first review evidence for modality specific temporal processing, such as how one modality can affect the other. Asymmetries in the potential of one modality to bias are another supports distinct temporal processing across modalities. As we will see, a range

of findings report auditory dominance for multi-modal sensory information in the time domain.

1.4.1.1 Audition and Vision

An auditory dominance over vision in the time domain is generally reported, hypothesised to be caused by greater temporal sensitivity in audition. This dominance is for instance observed when conflicting information is provided in the two modalities (Recanzone, 2003; Repp & Penel, 2002; 2004). The phenomenon, often referred to as "temporal ventriloquism", has been reported from various types of tasks. For example, the reproduction of temporal stimuli by a button press is known to be superior in audition than vision, as less variability is observed for auditory stimuli (Glenberg & Jona, 1991; Glenberg, Mann, Altman, Forman, & Procise, 1989). The auditory superiority can also be seen on temporal order judgment tasks (Morein-Zamir, Soto-Faraco, & Kingstone, 2003), as well as temporal localisation task: to judge whether a second sensory signal was presented closer to a first signal or a third signal, for a group of three sensory signals (Burr, Banks, & Morrone, 2009). Burr, Banks, and Morrone (2009) found a better temporal localisation ability for audio-visual stimuli than for unimodal stimuli, which can be explained by the theory of "optimal cue combination": if timing information is more accurate in audition, auditory cues would dominate timing-related tasks.

Timing perception can be disrupted by this modality difference. A series of studies found modality effects on time related tasks. Burr et al. (2009) reported that temporal asynchronies were more difficult to detect for cross-modal stimuli compared to unimodal stimuli. This suggests the existence of mechanisms in individual modality that may not necessarily integrate fully for temporal judgments.

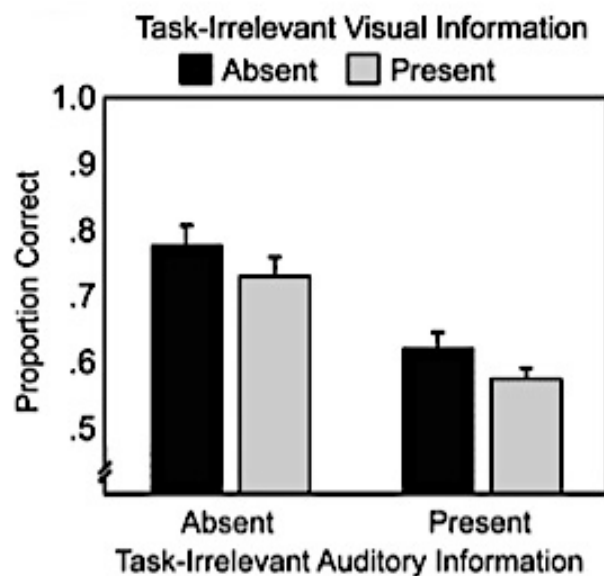


Figure 1-15. Results of a temporal discrimination task in the visual domain in the presence of task-irrelevant information. Visual performance level was mildly disrupted by the presentation of an incongruent visual time pattern (black bars vs. grey bars). The presentation of an incongruent auditory time pattern showed much greater disruption on the discrimination task performance (left black and grey bars vs. right black and grey bars). From Guttman et al. (2005).

Guttman, Gilroy, and Blake (2005) also found that the comparison of two visual streams could be disrupted when an incongruent, task irrelevant auditory stream was presented. Guttman et al. presented auditory and visual stimuli either with consistent or inconsistent during a temporal sequence discrimination task and asked participants to focus on only one modality. Performance was better in the congruent multisensory condition compared to the condition with only visual stimuli. Furthermore, a presentation of task-irrelevant auditory stimuli (incongruent auditory temporal patterns) highly disrupted performance in the visual temporal discrimination task, more so than the presentation of task irrelevant visual information (Figure 1-15). Based on these findings, they suggested that temporal sequences processing recruited some form of auditory encoding regardless of the modality of entry. In the example they studied, they argued that visual temporal patterns were coded in an auditory format for the sequence discrimination task.

1.4.1.2 Audition and Touch

A close inter-connection between audition and touch for temporal processing has been frequently reported (Mahar, Mackenzie, & McNicol, 1994; Yau, Olenczak, Dammann, & Bensmaia, 2009). Mahar and colleagues (1994) suggested that both the auditory and tactile modalities were designed to efficiently process temporally distributed information for threshold tasks, whereas visual modality was more adapted to information processing with spatial (or spatio-temporal) information.

The study of Ley, Haggard, and Yarrow (2009) examined whether more accurate temporal judgments could be performed when timing information was introduced in multiple modalities or not. They used a temporal order judgment task asking participants to judge the order of two stimuli when stimuli were presented one on the left and the other on the right side (auditory stimuli over different loudspeakers and tactile stimuli over different fingers). Stimuli were presented either unimodally or bimodally. They reported better performance in the temporal order judgment task when auditory and tactile stimuli were presented bimodally, compared to when information was presented in a single-modality. This indicates that participants were able to integrate stimuli, rather than relying one modality, to distinguish temporal events.

There is also evidence of auditory dominance in audio-tactile task, comparable to what we have seen previously for vision. The study of Roy, Dalla Bella, and Jagarde (2017) asked participants to flex their index finger as soon as a stimulus was presented. First, they showed that the response time to tactile only stimulus was slower than to auditory or to synchronised audio-tactile stimuli. Subjects were further asked to synchronise to audio-tactile stimuli presented with different stimulus-onset-asynchronies (SOAs) by focusing on either modality. The best synchronisation was observed when tactile stimuli preceded auditory modality by 40 and 80 ms. When subjects were asked to focus on auditory stimuli, the synchronisation performance was not different compared to performance with audio-only patterns. However, bigger variances to the synchronisation and coordination were found when subjects were asked to focus on tactile stimuli for audio-tactile patterns with SOAs, again compared to performance for tactile only patterns (Figure 1-16). The results indicate a strong auditory dominance for processing multi-modal temporal patterns over the tactile modality.

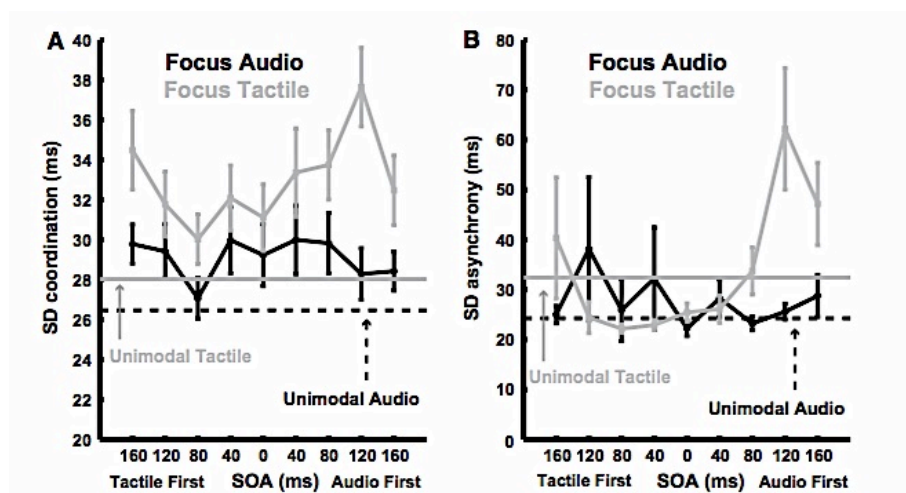


Figure 1-16. Results of a synchronised tapping task to audio-tactile stimuli. Average standard deviation (SD) on bimanual coordination between two hands (left) and asynchrony (right) with a focus on one modality when two modalities were presented with different SOAs. Variance on the auditory tapping task did not differ between audio-only temporal pattern and audio-tactile with SOAs. On the other hand, variance on the tactile tapping task increased when tactile stimuli were simultaneously presented with auditory stimuli with SOAs. From Roy et al. (2017).

1.4.1.3 Further studies across modalities

Some studies, although relatively few, have reported distinct temporal processing mechanisms across sensory modalities by comparing behavioural differences between vision and touch, or between audition, vision, and touch together. Grondin & Rousseau (1991) used an interval judgment task with empty intervals in all three sensory modalities, using a single-stimulus method where subjects were asked to determine whether an interval was short or long. Interval markers were either intermodal, across audition-vision-touch, or intramodal. For some trials, participants were not informed about the modality of the interval. Performance was better for the intra-modal presentation compared to inter-modal presentation, but no difference was observed between certainty and uncertainty condition. They suggested that the better performance for the intra-modal stimuli than the inter-modal stimuli was due to the asymmetry on temporal processing of each modality rather than the certainty of attention to the stimuli.

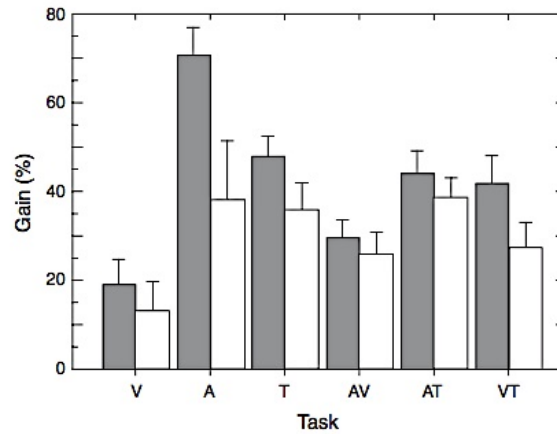


Figure 1-17. Results of a simultaneity judgment task between two temporal patterns, in different modalities. After 7 days of practice and test session, the performance gain (test performance - practice performance) on the 8th day (grey) and the gain from a re-test session after about 7 months (white) are plotted. The gain was always smaller when two patterns were presented in different modalities. From Virsu et al. (2008).

Additionally, Virsu, Oksanen-Hennah, Vedenpää, Jaatinen, and Lahti-Nuuttila (2008) used a simultaneity task asking whether two series of pulses were simultaneous or not, with series presented either inter- or intra-modally. For the intra-modal presentation for each modality, they used different strategies to present stimuli. In the auditory modality (A), authors asked participants whether there were two stimuli (simultaneous) or three (non-simultaneous) stimuli, presented from the loudspeaker. In the visual modality (V), there were two LEDs that would either emit three flashes either in a simultaneous or non-simultaneous manner. Lastly in the tactile modality (T), a series of tactile stimuli would be presented to two fingers (index and middle) for the simultaneity judgement. For the inter-modal presentation, the stimuli were presented in two chosen modalities, either as audio (via the speaker)-visual (via one LED; AV), audio-tactile (on one fingertip; AT), or visuo-tactile (VT) combination. They trained their participants on the task and measured a performance gain after training. The gain was the smallest for AV combination, followed by VT, and AT. The learning of AT was also not as high as for the intramodal (T) condition (Figure 1-17). Significant differences on the task performance between inter- and intra-modal conditions as well as across modalities indicate that temporal processing could highly vary based on the modality.

Overall, distinct abilities for temporal processing across modalities are most frequently suggested by the auditory dominance over other two sensory modalities and

INTRODUCTION

disrupted performance for the multi-modal presentation of temporal information, compared to the uni-modal presentation.

1.4.2 Centralised mechanisms from cross-modal studies

So far, we have seen modality differences in temporal processing that can affect temporal perception of other modalities when presented together, generally with auditory dominance found. However, to decode multi-sensory input in an ecological and efficient way, sensory systems could also share common mechanisms. This has also been investigated in the context of temporal processing.

As we have seen, temporal resolution differs across sensory modalities. However, it could be that similarities appear across modalities after these differences in acuity are taken into account. The study of Marks (1987) explored this idea by comparing three modalities for temporal pattern perception above the limit of acuity. Temporal patterns were composed of three successive pulses (of pulse duration either 100 ms or 200 ms) which thus produced two intervals, one long (either 450 ms or 650 ms) and one short (either 150 ms or 250 ms). Relatively big intervals used even for the short interval category helped minimising a modality effect from differences in temporal acuity. He introduced a pair of two temporal patterns, and asked subjects to judge the subjective similarity between two patterns. The pair of patterns was presented either in the same modality or in different modalities. The result showed that the cross-modal temporal pattern comparison was less sensitive to interval durations than the unimodal temporal pattern comparison. However, subjects' judgements were generally consistent for all unimodal and cross-modal presentations. This suggests similar perceptual dimensions for timing judgments in all three modalities when temporal acuity is not a limiting factor.

Rammsayer (2014) also pointed out that the contradictory results on temporal processing across modalities may be due to the wide variability of tasks and stimulus structure across studies. He used various structural combinations to test discrimination abilities between standard and comparison intervals, as a function of the type of interval (filled or empty), task (reminder task or 2AFC task), the sensory modality, and the base duration for standard stimuli.

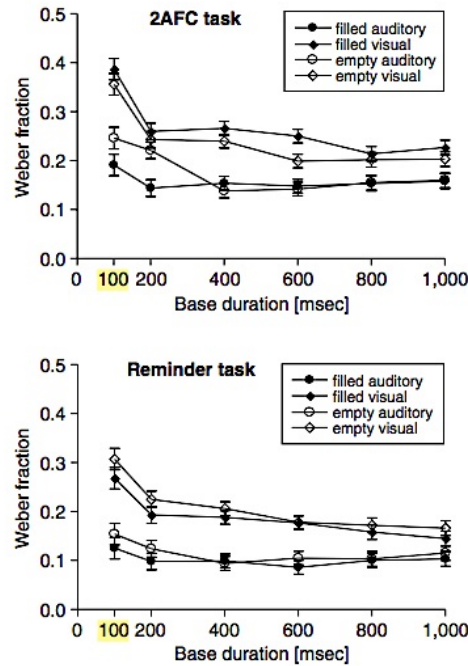


Figure 1-18. Results of interval discrimination task indicated by Weber fraction ratios as a function of interval type (filled vs. empty), task type (2AFC vs. Reminder), modality (auditory vs. visual) and the base duration (100 ms to 1000 ms). The reliable effect of modality task type, and base duration on values of the Weber fraction was found. From Rammsyer (2014).

Rammsayer found differences in performance depending on all experimental parameters. Generally, performance on the interval discrimination task was better in audition compared to vision. Also, performance was worse when the base duration was relatively short (100 – 200 ms) compared to longer durations, regardless of the task and modality. The modality effect started to diminish for base duration longer than 600 ms (Figure 1-18). The results indicate a possibility of observing similar performance of time-related behavioural tasks, potentially using the same strategies of temporal information processing, when the task and stimuli are carefully controlled.

1.5 The aims of the thesis

This introduction has only surveyed a small fraction of the abundant literature on timing perception. Even so, it already appears that there are important gaps in our knowledge of temporal information processing. The present thesis focuses on one of those gaps: are there mechanisms in audition and other modalities to rapidly learn temporal patterns, with

INTRODUCTION

patterns more complex than single intervals and without the special case of rhythmic information. We will use a novel variant of the experimental paradigm of Agus et al. (2010), using random time patterns as stimuli. The first main experimental chapter describes this novel paradigm and applies it to auditory time patterns (Chapter 2). Because time patterns are relevant for all modalities tested, in the second main experimental chapter we then transpose the paradigm to touch and vision, being careful to use intervals beyond the temporal acuity in all modalities for a fair comparison (Chapter 3). Two additional chapters, with more preliminary data and analyses, will introduce further findings regarding the learning of auditory time patterns in electric hearing with a cochlear implant, as well as the use of pupillometry to track the learning process. An appendix will finally be presented illustrating ongoing experiments using electrophysiological recordings.

The thesis will be concluded with a final chapter (Chapter 6) summarising all of the presented projects with discussions on future perspectives for understanding temporal perception, and its applications in the field.

Chapter 2

Auditory memory for random time patterns

This chapter has been published as:

Kang, H., Agus, T. R., & Pressnitzer, D. (2017). Auditory memory for random time patterns. *The Journal of the Acoustical Society of America*, 142, 2219-2232.

Parts of this work have been presented as:

Kang, H., Agus, T. R., & Pressnitzer, D. (2015). Auditory Memory for Time-Domain Information. 38th MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, USA (*Talk*).

2.1 Abstract

We investigated the acquisition of auditory memory for temporal patterns. The temporal patterns were random sequences of irregularly spaced clicks. Participants performed a task previously used to study auditory memory for noise (Agus et al., 2010). The memory for temporal patterns displayed strong similarities with the memory for noise: temporal patterns were learnt rapidly, in an unsupervised manner, and could be distinguished from statistically matched patterns after learning. There was, however, a qualitative difference with the memory for noise. For temporal patterns, no memory transfer was observed after time reversals, showing that both the time intervals and their order were represented in memory. Remarkably, learning was observed over a broad range of time scales, which encompassed rhythm-like and buzz-like temporal patterns. Temporal patterns present specific challenges to the neural mechanisms of plasticity, because the information to be learnt is distributed over time. Nevertheless, the present data show that the acquisition of novel auditory memories can be as efficient for temporal patterns as for sounds containing additional spectral and spectro-temporal cues, such as noise. This suggests that the rapid formation of memory traces may be a general by-product of repeated auditory exposure.

2.2 Introduction

Listening often requires matching current sensory evidence with stored memory traces, such as when associating a series of artificial tones to one's new phone (Royer et al., 2010), or recognizing natural sounds like voices and musical instruments (Agus, Suied, Thorpe, & Pressnitzer, 2012; Norman-Haignere et al., 2015). In fact, as auditory information develops over time, most aspects of auditory perception are likely to be shaped by memory processes on different time scales (Demany & Semal, 2007). How such memories are acquired through experience is not well understood. Here, we investigated auditory memory for temporal patterns, using random sequences of irregularly spaced clicks, over a broad range of time scales regarding the time-intervals forming the patterns.

In previous studies probing the acquisition of novel auditory memories, another type of random stimulus has been extensively used—white noise, obtained by drawing successive audio samples from a Gaussian distribution (Agus et al., 2010; Viswanathan et al., 2016). Studies of the memory for noise differed from each other in many experimental details, but, essentially, an exemplar of noise selected at random was presented to listeners more than once during an experiment. Putative memory traces were then probed by tracking psychophysical or physiological measures for the re-occurring noise exemplar, as exposure increased, either within a trial or across trials. These studies have demonstrated remarkable features of auditory memory. Immediate auditory memory for noise, often probed by asking participants to detect ongoing repetitions of a “frozen” noise exemplar, can extend up to noise durations of several seconds (Guttman & Julesz, 1963; Kaernbach, 2004; Warren et al., 2001). Longer-term memory traces can be measured by presenting the same exemplar over different trials or experimental blocks, and in this case memory be established within tens of seconds and persist for minutes or even weeks (Agus et al., 2010; Viswanathan et al., 2016). Real-time physiological correlates of the memory traces can be observed in electroencephalography or magnetoencephalography, with a response timing and topography suggesting an origin within auditory cortex (Andrillon et al., 2015; Luo et al., 2013). Finally, these physiological correlates can be formed incidentally, even with a task diverting attention away from the noises to be remembered (Andrillon et al., 2015), or during REM sleep (Andrillon et al., 2017).

A puzzling question has been raised by most of these studies: what had been learnt by listeners when they displayed evidence of a memory for noise? Answering this question has implications for whether the memory for noise is representative of the memory for sounds in general. It is implausible that listeners learnt the exact time series of the thousands of audio samples comprising the noise waveform, first because of the sheer amount of information involved, and second because audio waveforms are transformed as they ascend the auditory system so the central nervous system does not have access to the audio samples themselves. Rather, a memory for noise should rely on a subset of auditory cues emerging after processing along the auditory pathways. Such cues can belong to two broad classes: spectral cues, like the random distribution of frequencies over a short time frame, or temporal cues, like the occurrence of salient “events” at random times during the stimulus. Even though the opposition between spectral and temporal cues is somewhat arbitrary at the audio signal level, being essentially imposed by the time frame of analysis, the distinction is meaningful and important from an auditory perspective—and even more so from an auditory *memory* perspective. To simplify, spectral cues may be encoded by the place pattern of spiking rates in frequency-tuned populations of auditory neurons, whereas temporal cues may be encoded by the timing between spikes within a neuron or across neurons in a population. Even though the mapping between acoustic cues and their underlying neural representation is not well-established (Lim, Lagoy, Shinn-Cunningham, & Gardner, 2016), both place (Rothschild et al., 2010) and temporal (Lu, Liang, & Wang, 2001) cues are abundant along the auditory pathways, from the auditory periphery up to at least the auditory cortex. However, the two types of cue require qualitatively different mechanisms when learning and neural plasticity are considered. Spectral cues in a place code are available as simultaneously active neurons, so the classic “fire together, wire together” principle of neural plasticity can apply (Masquelier, Guyonneau, & Thorpe, 2008). For temporal cues, however, such a mechanism does not apply directly because what needs to be learnt develops over time. So, neurons would need somehow to encode past and present information simultaneously to learn temporal patterns (Lim, Lagoy, Shinn-Cunningham, & Gardner, 2016).

For noise, the evidence available so far points strongly towards a memory representation based on brief spectral cues. In their original report, Guttman and Julesz (1963) showed that listeners could detect repeating copies of a noise segment even for periodicities outside of the pitch range. They went on to describe the perceptual cues used

to detect repetitions with noise: a series of brief events, such as “rasps” and “clangs”, which led them to the following statement: “*It appears to us that the basis of pitchless periodicity detection must lie in the detection of short-term power-spectrum recurrence.*” (Guttman & Julesz, 1963). This conjecture, based on introspection, was confirmed experimentally by Kaernbach (1993), who used a technique based on participants tapping to a repeated noise exemplar. The replacement of various portions of the memorized noise exemplar with fresh noise showed that repetition-detection was based on short segments, lasting less than 100 ms. Further experiments using behavioral reverse-correlation suggested that, while some of those short features could be described as spectro-temporal, most were spectral edges (Kaernbach, 1999; 2000; and pers. comm.). More recently, when studying longer-term memory for noise, Agus et al. (2010) showed that the memory representation was robust to the time reversal of the noise segment. Time-reversal would leave short-term power spectra unaffected, but would disrupt most cues based on timing. Viswanathan et al. (2016) reported similar conclusions using time shuffling instead of time reversal. Finally, Andrillon et al. (2015) observed sizeable EEG event-related potentials after learning. As event-related potentials require a precisely time-locked neural response, this finding was accounted for by positing neural responses locked to brief spectral features within the noises after learning.

Considering all of these pieces of indirect but converging evidence, together with the well-established availability of neural mechanisms to learn spectral cues, it is legitimate to ask whether the remarkable characteristics of auditory memory observed with noise are restricted to sounds containing spectral cues. Here we aimed to answer this question directly by using random time patterns devoid of spectral cues.

The perception of auditory time patterns has been studied extensively, mostly using discrimination experiments. The simplest possible temporal pattern is a single time interval of variable duration. The discrimination between two such intervals yields thresholds corresponding to a Weber fraction, of about 10% for intervals in the range of about 200 ms to 2 s—although this range may vary with experimental details (Grondin, 2010). When the time interval to be discriminated was embedded within longer, isochronic patterns, or relatively simple patterns containing only two different interval durations, the Weber law held (Hirsh et al., 1990). However, when the embedding patterns were more complex, other listening strategies appeared. Using a discrimination task between two fully random temporal patterns, Sorkin (1990) showed that behavioural results could be accounted for by a model based on the cross-correlation between time

intervals, each interval being corrupted by additive noise (which would be inconsistent with a Weber fraction). When the embedding pseudo-random patterns were manipulated to induce a musical meter of varying strength (Povel & Essens, 1985), the detailed structure of the sequence influenced discrimination performance, demonstrating that not all intervals had the same perceptual weight (Ross & Houtsma, 1994; Teki & Griffiths, 2014). Finally, in yet another departure from the Weber law, when the interval to be discriminated was preceded by another, random time interval, this systematically distorted the perceived duration of the target interval (Karmarkar & Buonomano, 2007).

A comprehensive theory accounting for all of these findings about temporal pattern discrimination is still lacking, but, as outlined by Ross and Houtsma (1994), it may be useful to consider at least three types of strategy to discriminate complex time patterns: an analytic strategy, focusing on each interval independently (Hirsh et al., 1990); a holistic strategy, considering the whole pattern (Sorkin, 1990); and an emphasis on a small subset of pattern motifs (sometimes termed “runs” as they occur in time), which may stand out because of meter (Povel & Essens, 1985), perceptual organization (Royer & Garner, 1966), or informational content (Pollack, 1968a). Similar strategies could be relevant for the memory representation of time patterns, but this has not been investigated yet.

In the present set of experiments, we used irregularly spaced click trains in a memory paradigm similar to what has been used previously for noise (Agus et al., 2010). The intervals between clicks were randomly derived from a uniform (Experiment 1) or Poisson distribution (Experiment 2), to evaluate the influence of interval statistics on performance. A minimum gap between clicks was always imposed, as well as high-pass filtering, to prevent any overlap in the internal representation of successive clicks and thus restrict auditory cues to purely timing cues (Methods, see also Figure 2-1). One condition consisted of a random series of clicks, lasting 2 s in total, which we termed *C* (Figure 2-1, top). A second condition, termed repeated clicks or *RC*, consisted of a random series of clicks lasting 1s, immediately repeated for another 1s (Figure 2-1, bottom). The task assigned to participants for all experiments was to report whether a trial contained a repetition or not. There was a third condition, not mentioned to participants during the instruction phase: a Reference *RC*, or *RefRC*, which was an *RC* that re-occurred over several trials throughout an experimental block. As we showed previously, longer-term memory can be probed by comparing performance on *RCs*, heard in one and only one trial, to performance on *RefRCs*, heard identically over several trials (Agus et

al., 2010). Improved performance for RefRCs compared to RCs can be taken as an indication of longer-term memory traces. In Experiment 3, we tested if time-reversal would affect performance after learning, as a memory representation based on local cues such as extreme values of the inter-click intervals could be resilient to time-reversal. Finally, we used Sorkin's cross-correlation model (Sorkin, 1990) to see if it could help interpret some characteristics of repetition-detection (RC performance) for our click trains.

Unlike previous studies of auditory temporal patterns, we probed both the immediate memory processes needed for discrimination, reflected by the performance on RC trials, and longer-term memory processes over the course of minutes reflected by a possible advantage of RefRC trials over RC trials. The experimental design precisely controlled for the amount of exposure to RefRC, as participants would not have heard the exact same random time pattern before, so we could characterize the time-course of memory acquisition for temporal patterns. Finally, we could use the same psychophysical task to test a broad range of time scales (inter-click intervals), covering sequences sounding either like a succession of isolated clicks or sequences sounding like continuous buzzing sounds.

2.3 Experiment 1: uniformly distributed time patterns

A. Method

1. Participants

Fourteen participants were tested (age in year $M = 25$, $SD = 4$; 6 female). All had normal hearing, as established by an audiogram administered before the experiment (20 dB Hearing Level (HL) or less for all tested frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz). Participants gave informed consent and were paid 10 euros per hour. The study was approved by the ethical committee of U. Paris Descartes, France (CERES, IRB: 20154000001072).

EXPERIMENT 1: UNIFORMLY DISTRIBUTED TIME PATTERNS

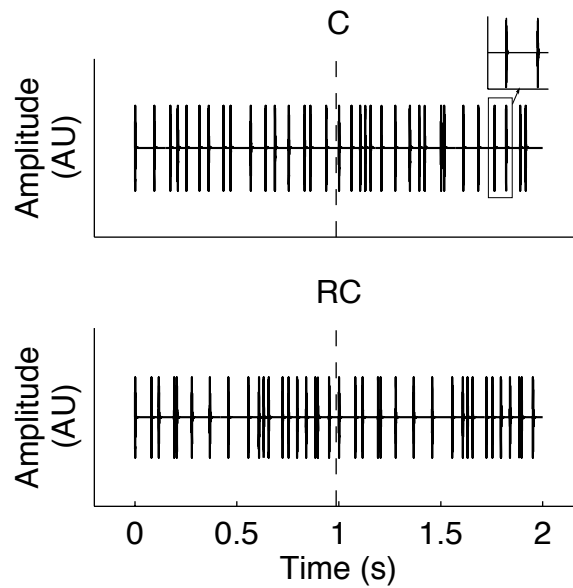


Figure 2-1. Illustration of the random auditory time patterns used in all experiments. Stimuli were 2 kHz high-pass filtered click trains, combined with low-pass noise, with a minimum time gap of 10 ms imposed between clicks. The condition illustrated corresponds to a uniform distribution between intervals and a 18-Hz average click rate, with the low-pass noise omitted for clarity. The output of a single auditory filter (gammatone) centred at 4 kHz is represented. The vertical dashed lines indicate the midpoint of the stimulus. On the top panel, a random pattern is shown (condition C, see text). On the bottom panel, the first and second halves of the stimulus are the same random pattern (conditions RC or RefRC). The inset on the top panel illustrates that the minimum gap between clicks ensured that there was no overlap between clicks, even after auditory filtering.

2. Stimuli

Click trains were generated in the digital domain at a sampling rate of 44.1 kHz. Each click was a rectangular pulse of two samples (about 50 μ s). Inter-click intervals (ICIs) were drawn from a uniform distribution, whose lower boundary was maintained constant at 10 ms for all experiments. The upper boundary of the ICI distribution was an experimental parameter. Five logarithmically spaced values were tested: 25, 50, 100, 200, and 400 ms. When taking account of the upper and lower boundaries, the five experimental conditions corresponded to average click rates of 57 Hz, 33 Hz, 18 Hz, 10 Hz, and 5 Hz. For consistency with Experiment 2, average click rate is used to identify conditions.

Auditory filter ringing may introduce overlap in the excitation caused by successive clicks, and thus spectral cues that can be potentially be used to discriminate stochastic click trains (Pollack, 1968b). To prevent the appearance of any such cues,

AUDITORY MEMORY FOR RANDOM TIME PATTERNS

all click trains were high-pass filtered using an 8th order Butterworth filter with a 2 kHz filter cut-off. The cut-off value was chosen as the 20th potential “harmonic” of the minimum ICI of 10 ms.

The nominal duration of each click train was 2 s. To form the first half of the stimulus, a first click was set at time 0 and successive ICIs were randomly drawn from a uniform distribution until the nominal duration of 1s was surpassed, and then the stimulus was truncated to 1 s minus the minimum gap of 10 ms. In the C condition, the procedure was repeated for the second half of the stimulus, with a fresh random draw so as to obtain a fully random click train lasting 2s (Figure 2-1, top). In the RC condition, the first half of the stimulus was repeated identically to obtain a repeated click train lasting 2 s (Figure 2-1, bottom). For C and RC, the random ICIs were generated anew for each trial. For RefRC, the same generation procedure as RC was adopted, but, additionally, ICIs were fixed for all trials within an experimental block. Note that the truncation at 1s caused a slight difference in the long-term distribution of the last ICI of the first half of the stimulus compared to all other ICIs. However, such a difference was only apparent on the long-term average of ICIs (each individual draw was compatible with the target distribution) and, in any case, it was present for all conditions.

Low-pass pink noise was finally added to the stimuli, to mask any possible distortion and to obtain stimuli lasting exactly the same duration. The noise pedestal was generated in the spectral domain with components between 50 Hz and 2 kHz. It started 200 ms before the click train and finished 200 ms after its nominal ending, including 50 ms raised cosine fade in and fade out. The root-mean square (*rms*) level of the pink noise was set at -20 dB relative to the click train *rms*. The full stimuli were presented at an overall level of 65 dB sound pressure level (SPL), A-weighted. As overall level was maintained constant, this means that peak level varied across (but not within) average rate conditions.

3. Apparatus and procedure

Participants were tested individually in a double-walled sound-insulated booth (Industrial Acoustics). Stimuli were played diotically through an RME Fireface UC soundcard, at a sampling rate of 44.1 kHz and 16-bit resolution. They were presented over Sennheiser HD 600 headphones. Participants were instructed to report within-trial repetitions. They responded through a computer keyboard in a self-paced manner.

Prior to data collection, a training period was included. Training was intended to help participants discover the types of perceptual cues elicited by repeated click trains. We modelled this training period on what had been previously used for repeated noise experiments (Agus, Carrión-Castillo, Pressnitzer, & Ramus, 2014). Training started with 10 repeats of 1-s long random click trains, a task in which informal listening indicated that repetition detection should be easy. The click trains were generated as Cs and RCs in the main experiment, with an average click rate of 12 Hz. This rate was chosen to be approximately in the middle of the experimental parameter range, but different from any subsequent experimental parameter. Participants received visual feedback on the repetition detection task during training. The first training block presented 10 trials, half of them C and the other half RC. Then, training blocks were run with 4, then 3, then 2 repeats only, for 40 trials in each training block. Note that there were no RefRCs in any of these training blocks.

When data collection began, no feedback was further provided. The instructions to participants remained the same. In particular, at no time were they instructed about the possible re-occurrence of a sound (RefRC) during the block, nor about our aim to test memory. In each experimental block, the experimental parameter (average click rate) was fixed and there were 40 C, 20 RC and 20 RefRC trials in pseudo-random order. The only constraint on order was that RefRCs could not appear on successive trials (Agus et al., 2010). Participants completed 4 blocks for each of the 5 click-rate conditions. Blocks for a given parameter were run in succession, but the order of parameters was random and counterbalanced across subjects. Training and testing was completed over two experimental sessions, each lasting about 2 hours and run on consecutive days.

4. Data analysis

The pattern of “yes” and “no” responses was first converted to hit rate (proportion of “yes” responses for RC or RefRC) and false alarms (“yes” response for C). The sensitivity index d' from signal detection theory was then computed for each participant and RC or RefRC conditions (sharing false alarms for a given experimental block). Statistical testing was performed on d' with repeated-measures ANOVAs, with trial type (RefRC or RC) and average click rate (5, 10, 18, 33, 57 Hz) as within-subject factors and participants as random factors. We verified that the data conformed to the assumptions required for ANOVA and applied Greenhouse-Geisser corrections when necessary. Along with the F statistics and p -values, with a

significant level set at $p < 0.05$, we also report effect size with the generalized eta-squared measure, η^2_g (Lakens, 2013). Post-hoc t -tests were conducted, corrected for multiple comparisons by using the Holm-Bonferroni method as necessary.

When analyzing the time course of performance, responses were converted to hit rate and aggregated in trial bins over all blocks and participants. The first bin represents the first time participants heard e.g. a RefRC during a block, but not necessarily the first trial of the block and not necessarily the same trial number for all participants. These average values were then fitted with either a flat line corresponding to the average performance (1 parameter) or exponential functions (3 parameters) using a least-squared method. The goodness of fit was estimated by penalizing models with extra parameters (Motulsky and Christopoulos, 2004). The choice of model then led us to conclude whether the performance of RC and RefRC evolved over time, perhaps because of learning, or stayed statistically constant over the duration of the blocks (Agus et al., 2010).

EXPERIMENT 1: UNIFORMLY DISTRIBUTED TIME PATTERNS

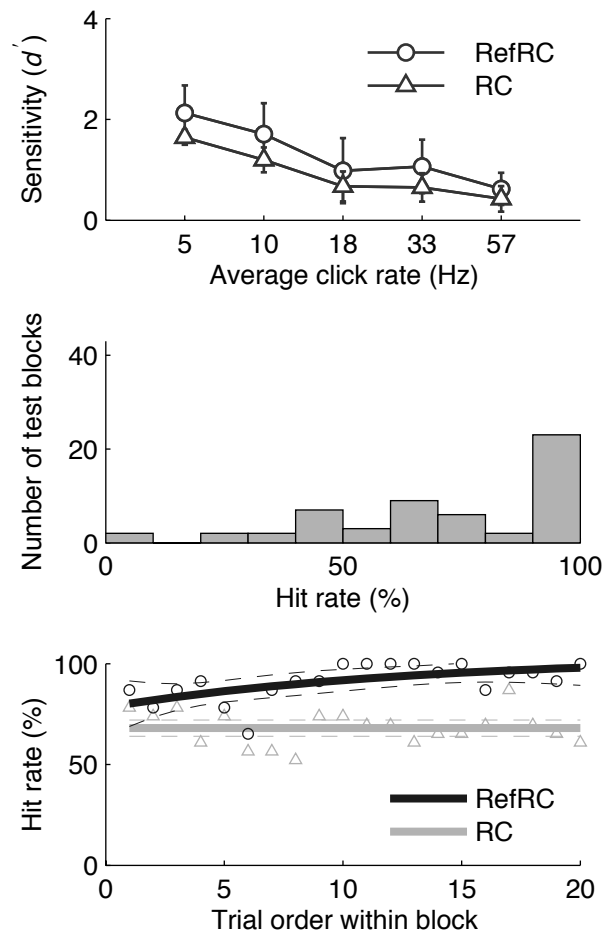


Figure 2-2. Results for Experiment 1, Uniform distribution. Top panel: The sensitivity index d' is shown for each stimulus type and average click rate conditions, averaged across listeners. Error bars are 95% confidence intervals about the mean. Middle panel: Histogram of the proportion of hit rates for the last 10 trials of each block in the RefRC condition, over all blocks for all listeners. Here, only the condition with 18 Hz average rate is shown, but all conditions displayed similar results. Lower panel: The time course of hit-rate for all blocks with 90% or more hit-rate over the last ten trials of a block. Data points are average values over blocks and listeners for each trial order. The solid lines are best-fit exponentials for each stimulus condition, with dashed lines indicating 95% confidence interval. Again, only the 18 Hz condition is illustrated¹.

B. Results

The top panel of Figure 2-2 displays the average performance across participants as the sensitivity index d' , for RC and RefRC trials, and over the five values of click

¹ see *supplementary information* for the time course of hit rates for all rate conditions.

rates tested. A first feature of the results is that performance generally decreased with increasing click rate, from d' values around 2 at the lowest rate of 5 Hz, to d' values below 1 at the fastest rate of 57 Hz. Stimulus duration was kept constant, so increasing rates also corresponded to more clicks and thus more time intervals in the patterns. A second feature is that performance was generally higher for RefRC than RC, with an advantage of about 0.5 d' units for all rates.

A repeated-measures ANOVA on d' (see Methods) confirmed that both trial type ($F(1, 13) = 41.56, p < 0.001, \eta^2_g = 0.05$) and click rate ($F(4, 52) = 25.29, p < 0.001, \eta^2_g = 0.27$) had a reliable effect on performance. No interaction reached our significance criterion (all $p > 0.05$). Post-hoc comparisons confirmed that the advantage of RefRC over RC was present for all rates (5 Hz: $t(13) = -3.10, p < 0.01$, 10 Hz: $t(13) = -3.86, p < 0.01$, 18 Hz: $t(13) = -2.22, p < 0.05$, 33 Hz: $t(13) = -3.45, p < 0.01$, 57 Hz: $t(13) = -2.84, p < 0.01$; reported p -values are uncorrected but all $p < 0.05$ after Holm-Bonferroni correction).

A superior performance for RefRC over RC could indicate learning of RefRC thanks to increasing exposure to the reference click train as the block progressed. But it could also simply be the chance result of the particular set of randomly chosen RefRCs being easier to detect from the outset. To test for this possibility, we examined the changes in performance over trials throughout the block. Previous experiments with white noise showed that not all blocks produced learning (Agus et al., 2010). Moreover, RefRC performance was associated to higher variance than RC in Figure 2-2, which would also suggest an additional source of variability for this condition. Therefore, we first tested for the presence of “good” blocks and “bad” blocks in the present data. “Good” blocks were defined as blocks for which performance was high for RefRCs at the end of the block. The middle panel of Figure 2-2 shows the histogram of hit rates for RefRCs over the last 10 presentations of RefRC during the blocks, illustrated here for the 18 Hz rate condition. A one-sample Kolmogorov-Smirnov test confirmed that hit rates did not follow a normal distribution ($D(56) = 0.55, p < 0.001$). This suggests that two types of blocks contributed to the histogram: “bad” blocks, with hits around 50%, and “good” blocks, with hits above 90%. We then selected those “good” blocks (41% of blocks in the 18 Hz rate condition) and computed performance for RC and RefRC for each trial position within a block. The result is shown in the lower panel of Figure 2-2. Performance was not

constant for RefRC: from the beginning to the end of the block, performance increased. In contrast, for RC there was a no change in performance.

The same qualitative description of the data was verified in all other conditions, and tested statistically. We first tested for the presence of good and bad blocks: all distributions deviated from normal (one-sample Kolmogorov-Smirnov test, 5 Hz: $D(56) = 0.95$; 10 Hz: $D(56) = 0.55$; 18 Hz: $D(56) = 0.55$; 33 Hz: $D(56) = 0.58$; 57 Hz: $D(56) = 1$, all $p < 0.001$). When applying the criterion of selecting blocks with hit rates above 90% for the last 10 presentations of RefRC, the proportion of good blocks was 64%, 55%, 41%, 39%, and 23% for rate conditions of 5, 10, 18, 33, and 57 Hz, respectively. We next tested whether performance changed over the course of an experimental block, for RC and RefRC trials, by fitting exponential functions testing whether this explained the data better than a flat line at the mean hit rate (see Methods). This test confirmed that performance increased over the course of a block for RefRC, for all click rate conditions except 10 Hz (5 Hz: $F(2, 17) = 12.70$, $p < 0.001$, 10 Hz: $F(2, 17) = 2.18$, $p = 0.143$, 18 Hz: $F(2, 17) = 4.73$, $p < 0.05$, 33 Hz: $F(2, 17) = 11.85$, $p < 0.001$, 57 Hz: $F(2, 17) = 8.19$, $p < 0.01$). The same test suggested that performance for RC did not change over the course of a block for any of the click rates (all $p > 0.05$). RefRCs regarded as “bad blocks” also showed no performance change of the course of a block in any of the click rate conditions (all $p > 0.05$).

C. Interim discussion

The results obtained so far are similar to what had been observed using white noise in the same experimental paradigm (Agus et al., 2010). Across-trial re-occurrence of a stochastic time pattern, defined by a series of clicks with random ICIs, led to improved performance on a within-trial repetition detection task.

This was not observed for all experimental blocks, but only on between 23% and 64% of all blocks, depending on the baseline difficulty of the within-trial repetition detection task. These numbers bracket what was observed with naïve listeners and 0.5s-long white noise, where the proportion of good blocks was about a third. It is possible that the actual click train characteristics systematically differed between good and bad blocks, even though their generative parameters were the same. To test for this possibility, we compared between good and bad blocks the average ICIs, the standard deviation of ICIs, and the presence of exceptionally long or short ICIs. The analysis is detailed in the Appendix. Overall, we did not find any systematic

difference between good and bad blocks for these acoustic characteristics. As each RefRC was unique to a block and a participant, it is not possible to ascertain whether good blocks were caused by intrinsically easier RefRC patterns, with acoustic characteristics remaining to be elucidated, or if good and bad blocks rather resulted from an idiosyncratic interaction between pattern and participant.

The rapid time-course of learning was also similar with Agus et al. (2010), as 90% performance or better was reached after only 10 presentations of the RefRC sequence. Looking at the time course of the good blocks, we could rule out chance selection of easier stimuli for RefRC, but rather attribute the RefRC advantage to true learning. Thus, stochastic time patterns can rapidly form memory traces with increasing exposure.

The observed effect of click rate, a parameter that could not be tested with white noise, was perhaps in part predictable: slower rates, associated with clicks that could easily be heard out individually, produced better performance compared to faster rates. However, the advantage of longer-term learning in the case of RefRC was always of about 0.5 of a d' unit, with no interaction between this RefRC advantage and click rate. This suggests that even though faster rates led to poorer baseline performance and also less good blocks, the benefit of longer-term learning for RefRC remained about the same.

It is possible that the other aspects of the ICI patterns besides mean rate could have impacted performance. For instance, as we kept stimulus duration constant, faster rates led to more clicks and thus to more time intervals to be learnt. Even though it is intuitively appealing that more intervals should induce poorer performance, this has not always been observed in discrimination tasks. When the variance of ICIs is low, for instance, performance actually improves with number of intervals (Pollack, 1968a). We therefore repeated the experiment with a different random distribution of ICIs to test for the influence of additional ICI statistics besides the average click rate.

2.4 Experiment 2: poisson-distributed time patterns

A. Method

1. *Participants*

Participants were 14 normal-hearing listeners, who had not participated in Experiment 1 (age in year $M = 24$, $SD = 3$; 11 female). They were recruited as in Experiment 1, and their audiogram was confirmed as normal using the same procedure.

2. *Stimuli*

Stimuli were click trains generated as for Experiment 1, with a single difference: the ICIs were now drawn from a Poisson distribution, instead of a uniform distribution. We imposed a refractory time of 10 ms on the Poisson distribution, and adjusted its rate parameter to have the same average click rate conditions as for Experiment 1. To obtain average click rates of 5, 10, 18, 33, and 57 Hz, we used Poisson rate parameters of 5, 10, 22, 50, and 133 Hz, respectively. Stimuli were then high-pass filtered at 2 kHz and noise was added, exactly as described for Experiment 1. We again used conditions C, RC, and RefRC.

3. *Apparatus, procedure, and data analysis*

Apparatus, procedure, and data analysis were identical to Experiment 1.

AUDITORY MEMORY FOR RANDOM TIME PATTERNS

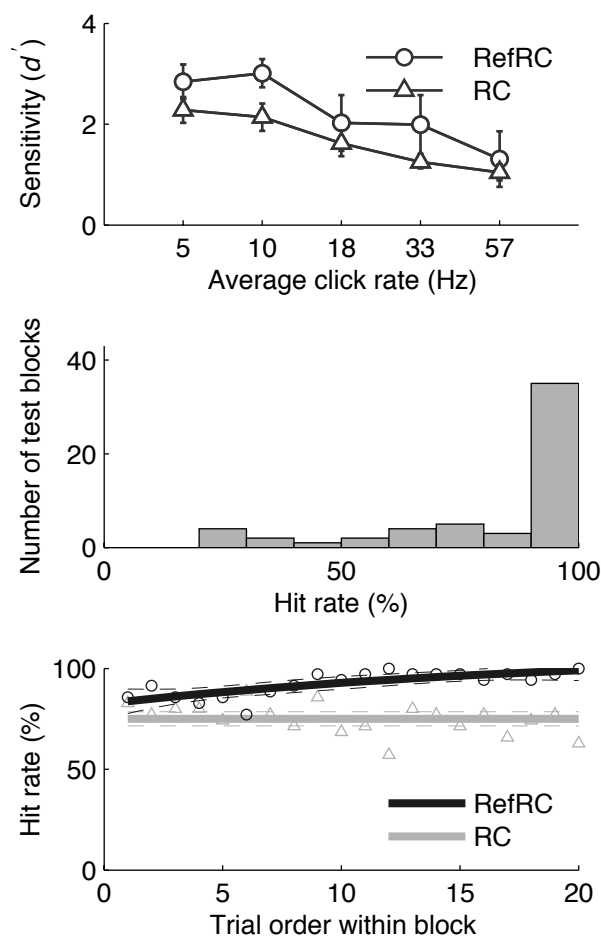


Figure 2-3. Results for Experiment 2, Poisson distribution. Same format as Figure 2-2.

B. Results

The top panel of Figure 2-3 displays the average performance across participants, for RC and RefRC trials, and for the five values of click rates tested. Results were broadly similar to those of Experiment 1: performance decreased with average click rates, and RefRC produced generally better performance than RC. However, the overall performance was higher here than for Experiment 1, with d' values between 2 and 3 for the slowest rates, and d' values above 1 for the fastest rates. The difference between RefRC and RC was in the same range or larger as what was observed for Experiment 1, between 0.5 d' units (5 Hz, 18 Hz, 57 Hz) and 1 d' unit (10 Hz, 33 Hz).

A repeated measures ANOVA was used to test those observations (see Methods, degrees of freedom corrected as appropriate). Both average click rate ($F(2.46, 31.93) = 25.25, p < 0.001, \eta^2_g = 0.38$) and trial type ($F(1, 13) = 82.40, p < 0.001, \eta^2_g = 0.14$) had a reliable effect on performance. An interaction with small

effect size was found between average click rate and trial type ($F(4, 52) = 3.02, p < 0.05, \eta^2_g = 0.02$). Post-hoc t -test confirmed that RefRCs produced higher performance than RCs for all average click rates (5 Hz: $t(13) = -2.42, p < 0.05$, 10 Hz: $t(13) = -4.20, p < 0.01$, 18 Hz: $t(13) = -2.59, p < 0.05$, 33 Hz: $t(13) = -6.92, p < 0.001, t(13) = -4.39, p < 0.001$; all $p < 0.05$ after the Holm-Bonferroni correction). The interaction may thus have been caused by an especially large performance advantage of RefRC over RC for 10 Hz and 33 Hz. Given the small effect size of the interaction, we did not explore it any further.

The middle panel of Figure 2-3 illustrates performance for RefRCs for different blocks in the 18 Hz average click rate condition. As for Experiment 1, the distribution of hit rate over the last 10 trials for RefRC was not normally distributed. This was confirmed by one-sample Kolmogorov-Smirnov tests for all rate conditions (5 Hz: $D(56) = 0.64, p < 0.001$, 10 Hz: $D(56) = 0.73, p < 0.001$, 18 Hz: $D(56) = 0.58, p < 0.001$, 33 Hz: $D(56) = 0.57, p < 0.001$, 57 Hz: $D(56) = 0.50, p < 0.001$). When applying the criterion of hit rate above 90% for the last 10 trials, we observed 79, 91, 63, 57, 38% of good blocks for rate conditions of 5 Hz, 10 Hz, 18 Hz, 33 Hz, and 57 Hz. The lower panel of Figure 2-3 illustrates performance for RC and RefRC during the course of good blocks. As for Experiment 1, performance increased for RefRCs while there was a small trend for a decrease for RC. Statistical analyses of the exponential fits showed that performance increased for RefRC during a block (5 Hz: $F(2, 17) = 20.80, p < 0.001$, 10 Hz: $F(2, 17) = 14.25, p < 0.001$, 18 Hz: $F(2, 17) = 11.21, p < 0.001$, 33 Hz: $F(2, 17) = 7.95, p < 0.01$, 57 Hz: $F(2, 17) = 13.64, p < 0.001$). For RC, performance decreased for 10 Hz ($F(2, 17) = 4.99, p < 0.05$) but was constant for all other rates (all $p > 0.05$).

C. Interim discussion

Experiments 1 and 2 used the same psychophysical paradigm and average stimulus characteristics, but with uniformly distributed (Experiment 1) or Poisson-distributed (Experiment 2) stochastic time intervals. The average performance and proportion of good blocks differed between the two experiments, with a general advantage to the Poisson distribution. We thus decided to compare formally the results of these two experiments.

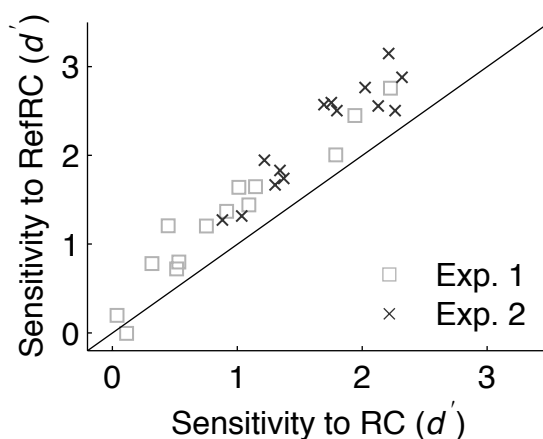


Figure 2-4. Comparison between Experiments 1 and 2. The sensitivity index d' is plotted for individual participants (squares: Experiment 1, crosses: Experiment 2) for RC and RefRC conditions. In all but one cases, the points are above the diagonal, indicating a performance advantage for RefRC.

Figure 2-4 shows the individual data obtained by all participants of Experiments 1 and 2, averaged across all blocks (good and bad) and averaged over all average click rates for each participant. A first observation, consistent with all analyses presented so far, is that in all cases but one (out of 28 participants), performance for RefRC was better than performance for RC.

It also appears that, in spite of individual variability, performance was generally better for Experiment 2. This was tested statistically with a mixed-design ANOVA on d' averaged across rates, with factors of experiment (1 or 2) and trial type (RC or RefRC). Both experiment ($F(1,26) = 12.37, p < 0.01, \eta^2_g = 0.09$) and trial type ($F(1,26) = 121.42, p < 0.001, \eta^2_g = 0.23$) had sizeable effects. A small interaction was also found ($F(1,26) = 4.52, p < 0.05, \eta^2_g = 0.004$).

Both experiments therefore clearly showed an advantage of RefRC over RC, but performance was generally higher for Poisson-distributed intervals compared to uniformly distributed intervals. The means of both distributions were matched across experiments, but distributions still differed in terms of higher order statistics. In particular, the Poisson distribution had higher variance. Using 1000 simulated click trains per condition, we estimated that the variance between successive time intervals was between 18% and 70% higher for Poisson-distributed stimuli, depending on average rate. As suggested before, higher variance may facilitate the discrimination between temporal patterns (Sorkin, 1990) and thus explain the difference between

Experiments 1 and 2. We evaluate this interpretation quantitatively in the modeling section below.

Another difference between Experiments 1 and 2 could have been in the local distribution of clicks. It is possible that salient “motifs” or “runs” were used by participants to perform the task (Pollack, 1968a; Ross & Houtsma, 1994). For instance, a chance succession of very short or very long ICIs could have been sufficiently salient to support the repetition detection task, even if the whole time pattern had not been learnt. Such runs may have occurred more often for Poisson distributions because of its higher variance. These runs are still temporal cues, but highly local.

To test for the importance of local cues on the results presented so far, we quantified the possibility of transferring the learning of one sequence to its time-reversed counterpart. If the whole time pattern was learnt, time reversal should dramatically impair performance. In contrast, if only local runs were learnt, such as distinctively long or short ICIs, or repetitions of identical ICIs, memory should transfer almost perfectly to time-reversed sequences. The same reasoning was followed by Agus et al. (2010) for white noise. They showed an almost-perfect memory transfer after time reversal for noise stimuli, suggesting local spectral cues for the memory representation of noise. In Experiment 3, we tested whether the same would hold for click trains.

2.5 Experiment 3: memory transfer after time reversal

A. Method

1. Participants

Twelve participants who had previously participated in psychophysical experiments in our laboratory were invited back as “experienced” participants (age in year $M = 25$, $SD = 3$; 5 female). We chose experienced participants because the experiment tested memory transfer, so for efficiency we wanted to ensure that as many blocks as possible would display initial learning. Participants’ audiograms were again verified as normal. All other details are as in Experiments 1 and 2.

2. Stimuli

The stimulus generation method of Experiment 2 was used, with Poisson-distributed ICIs. Here we only tested the average rate conditions of 5, 18, and 57 Hz. In addition to the C, RC, and RefRC trial types, a new “reversed-RefRC” trial type was introduced. This was simply achieved by reversing the order of ICIs of the RefRC trials (effectively, reversing the time axis of the stimulus before high-pass filtering was applied).

We used two minor variants of the stimulus generation procedure. In the method of Experiment 2, there was always a click at the beginning of the stimulus. As a consequence, all reversed-RefRCs would have displayed a click at the very end of the sequence, and thus a slightly different time-course relative to the noise pedestal compared to all other trial types. In a first variant of the stimuli, we shifted the reversed-RefRCs in time so that they would include a click at time 0s like all other conditions. In a second variant, we omitted the first click of the sequence for all trial types before reversal (C, RC, RefRC). This led to more variability overall in the start and end time of the stimuli relative to the noise pedestal, as well as a small change in average rate compared to Experiment 2, but here all conditions received the same treatment. Participants were randomly assigned to one or the other variant (6 participants for each).

3. Apparatus and procedure

The procedure was the same as Experiment 2, with minor differences. We used BeyerDynamics DT 770 Pro headphones. Also, as participants were experienced, no training session was provided.

The first half of each block contained 30 C, 15 RC, and 15 RefRC trials in pseudo-random order. The second half of the block seamlessly followed the first half without any interruption, and introduced the reversed-RefRCs. It contained another 30 C, 15 RC, and 15 reversed-RefRC trials. Four blocks were run in succession for each average rate. The order of presentation of the rates was random and counterbalanced across participants. Each experimental session lasted about 2 hours.

4. Data analysis

To test for the possible influence of the two possible variants in the stimulus generation procedure, the sensitivity index d' from signal detection theory was used to evaluate performance for RefRC and RC on the first half of the blocks. We then performed a repeated measures ANOVA on d' with factors of stimulus type, rate condition, and

stimulus generation procedure variant. To test for memory transfer, we estimated performance at various time points during the block by computing d' for subsets of trials. We first selected “good” blocks to ensure that the initial RefRCs had been learnt, using the criterion of 90% or more hit rate for the last 5 trials over which this stimulus was presented (which, for 5 trials, amounts to 100% hit rate). Then, for such good blocks, we computed d' for the RefRCs for the first 5 trials (i.e., before learning) and last 5 trials (i.e., after learning) of the first half of the block, estimating false alarms from the corresponding C trials. The same analysis was then carried out for reversed-RefRCs, for the second half of the selected blocks (i.e., when reversed-RefRCs replaced the original RefRCs). A perfect memory transfer from RefRC to reversed-RefRC would be reflected the same performance for the last 5 trials of RefRC and the first 5 trials of reversed-RefRC, before any re-learning could occur. To test for this, we used paired t -tests comparing d' on the last 5 trials for RefRC and first 5 trials for reversed-RefRC, for each rate condition. On some rates, one subject did not show any good blocks and was excluded from this particular analysis, hence the reduced degrees of freedom. All differences reported as statistically significant are at $p < 0.05$ after Holm-Bonferroni correction.

B. Results

We first confirmed that there was no effect of the stimulus generation variant on general task performance ($F(1,8) = 3.31, p > 0.05$) nor any interaction with other factors (all $p > 0.05$). All further analyses thus used aggregated data for the 12 participants, irrespective of stimulus generation variant.

A one-sample Kolmogorov-Smirnov test confirmed that hit rate for the last 5 trials of each block did not follow a normal distribution (5 Hz: $D(48) = 0.6423, p < 0.001$, 18 Hz: $D(48) = 0.5721, p < 0.001$, 57 Hz: $D(48) = 0.5929, p < 0.001$). We thus selected “good” blocks with hit rate of 100% for the last 5 trials, which led to selecting 67%, 70%, and 52% of blocks for 5, 18, and 57 Hz, respectively. For these blocks, the RefRC has been learnt, so we could meaningfully test for memory transfer to the reversed-RefRC.

AUDITORY MEMORY FOR RANDOM TIME PATTERNS

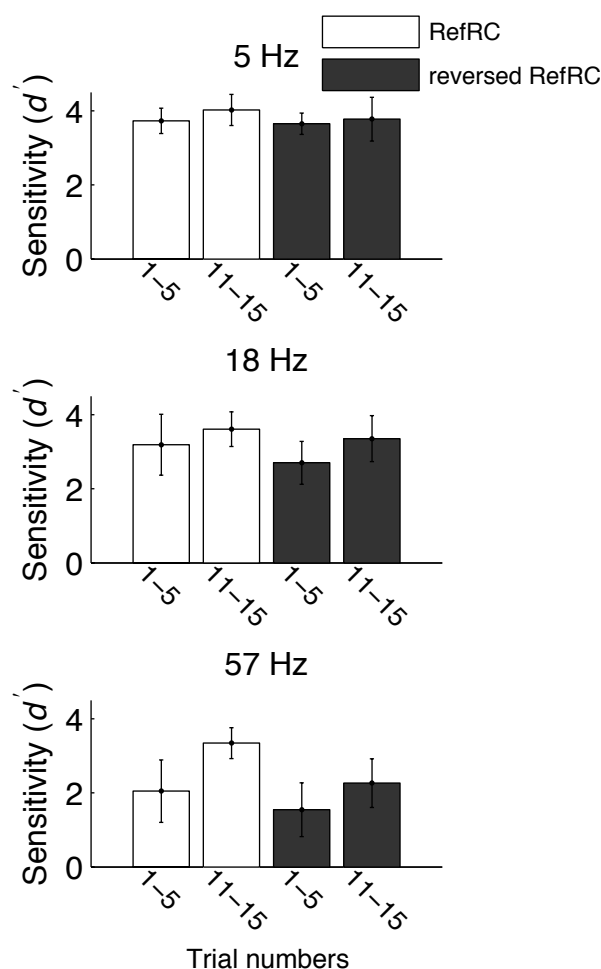


Figure 2-5. Results for Experiment 3, the effect of time-reversal. For each average click rate condition tested, performance is plotted at the beginning of the block (trials 1-5, white), just before the introduction of time reversals (trials 11-15, white), just after the introduction of time reversal (trials 1-5, black) and at the end of the block (trials 11-15, black). The effect of time reversal can be assessed by comparing the last white bar to the first black bar. Only blocks with 90% or more hit rate for the original RefRC are included. Results are averaged across listeners. Error bars are 95% confidence intervals about the mean.

Figure 2-5 displays the sensitivity to RefRCs and reversed-RefRCs over the course of a block. The important feature is what happened immediately after the introduction of time reversal (last white bar compared to first black bar). In all cases, performance dropped for reversed-RefRC, down to levels similar of those observed with initial exposure to a novel RefRC (first white bar). Paired *t*-test confirmed reliable performance drops after time reversal for rates of 18 Hz ($t(1,10) = 3.37, p < 0.001$) and 57

Hz ($t(1,10) = 6.96, p < 0.001$). The 5 Hz rate condition did not meet our significance criterion (5 Hz: $t(1,11) = 1.87, p = 0.11$).

B. Interim discussion

The pattern of results showed a drop in performance when a RefRC was replaced by its time-reversed version, with performance reverting to levels observed with novel RefRCs. The one condition that failed the significance test, 5 Hz average rate, did nevertheless exhibit the same pattern. In addition, this condition displayed equal initial performance for reversed-RefRC and RefRC, so the effect of time reversal was basically as large as it could have been. The lack of a statistically significant difference can thus be attributed to a ceiling effect, as d' values were high (above 4) at this slow rate with our group of experienced listeners.

Performance for the reversed-RefRCs followed a pattern compatible with the learning of a novel stimulus, with an increase in performance between the beginning and end of the second half of the block. The end performance for reversed RefRCs was generally not as high as that observed for the RefRC presented on the first half of the block, but that is likely because “good” blocks were selected based on performance for the initial RefRC only.

Overall, the data is consistent with little or no memory transfer from a learnt temporal pattern to its time-reversed version, which has to be re-learnt from scratch. This lack of memory transfer suggests that participants learnt extended temporal features, and not local runs of salient ICIs. This finding is perhaps especially intriguing for the high rate condition, 57 Hz. Such a high average rate is above the lower limit of pitch for isochronous click trains (Krumbholz et al., 2000; Pressnitzer et al., 2001). Accordingly, our stochastic sequences generally sounded like buzzes at this rate, with individual clicks difficult to hear out. However, even in this case, the memory representation appeared to retain an extended series of time intervals and their order of appearance.

2.6 Cross-correlation model

A. Rationale and model description

We adapted a model of sequence discrimination (Sorkin, 1990) to interpret the effect of ICI distributions on the repetition-detection task. The behavioural task that was assigned

to participants was to decide whether each trial contained a repetition or not. For modelling purposes, this can be construed as a discrimination task between the two halves, A and B, of the click train presented on any given trial. If A were judged sufficiently similar to B, the participant would respond “Yes” to the repetition detection task. For repeating trials, A and B were identical, so errors (misses) would be caused by internal noise. For non-repeating trials, A and B were different, so errors (false alarms) would occur if noisy representations of A and B were by chance as similar as noisy representations of identical stimuli.

This formalization of the task can be implemented using cross-correlation as a similarity measure between series of time intervals, corrupted by internal noise (Sorkin, 1990). The first half of the stimulus can be notated as $A = \{t_{A1}, t_{A2}, t_{A3}, \dots, t_{Am}\} + N(0, \sigma^2)$, with t_{Ai} the time interval corresponding to the i^{th} ICI and $N(0, \sigma)$ an additive noise term, following a normal distribution with mean 0 and variance σ^2 , added independently to each t_{Ai} . Similarly, the second half of the stimulus can be expressed as $B = \{t_{B1}, t_{B2}, t_{B3}, \dots, t_{Bn}\} + N(0, \sigma)$. To estimate the similarity between A and B, a cross-correlation was computed between the series $\{t_{A1}, t_{A2}, t_{A3}, \dots, t_{Ak}\} + N(0, \sigma^2)$ and $\{t_{B1}, t_{B2}, t_{B3}, \dots, t_{Bk}\} + N(0, \sigma^2)$, with $k = \min(m, n)$.

To compute d' values simulating our behavioural dataset, we ran the correlation model on ICIs derived from the experimental stimuli generation procedure, for the RC and C conditions. For each simulated trial, we selected a “Yes” response if the cross-correlation between A and B was above a fixed criterion, *Crit*. If the cross-correlation was below the criterion, a “No” response was selected. This provided a proportion of hits and false alarms for each trial type from which we could compute d' .

B. Results

After an initial exploration of the parameter range, the two free model parameters σ and *Crit* were systematically varied and the resulting goodness of fit of the model evaluated, using the squared error between behavioural data and model predictions. We computed d' values on 100 simulated C and RC trials, for all rate conditions and uniform or Poisson distributions, and for all tested σ and *Crit* values. The search space for σ was from 0 ms to 100 ms, in 5 ms steps, while for *Crit* it was from -0.9 to 0.9, in 0.1 steps.

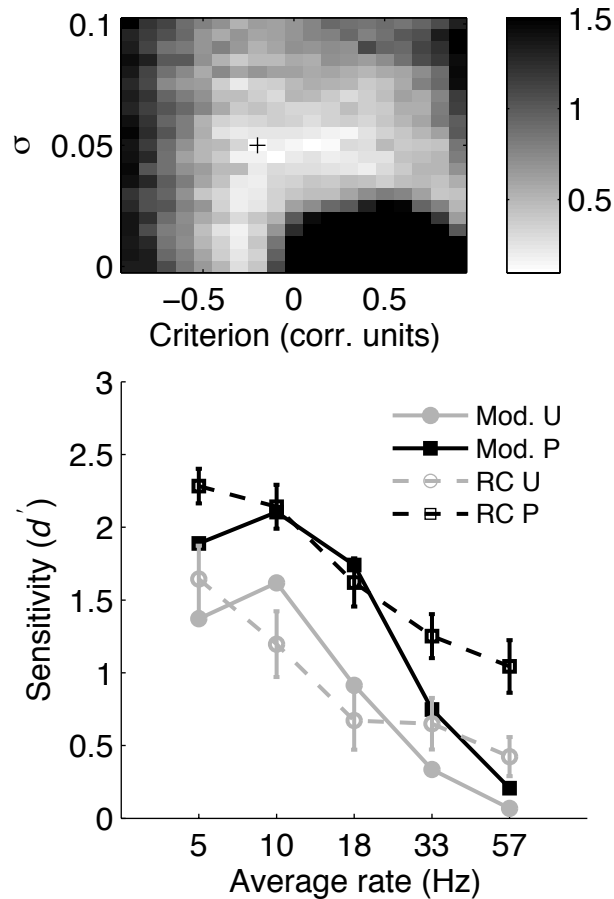


Figure 2-6. Prediction of a correlation model. Top panel: mean quadratic error relative to the behavioural data. The black cross indicates the best fit. Lower panel: the data for the RC condition are replotted from Experiment 1 (U) and 2 (P), along with the model predictions. See text for details.

A broad range of parameters led to a satisfactory fit to the data (Figure 2-6, top). The best fit was obtained for $\sigma = 50$ ms and $Crit = -0.2$. This estimate is different from the value of $\sigma = 15$ ms suggested in the original model (Sorkin, 1990). However, we verified that running the model with $\sigma = 15$ ms produced the same trends in the predictions, so the interpretations outlined below would hold with either value of σ .

For the fitted model parameters, 10000 simulated trials were computed to estimate d' . The model predictions are represented in Figure 2-6, bottom, along the corresponding behavioural data. The model correctly captured the two main trends in the data: first, lower click rates led to better performance; second, Poisson distribution led to better performance. However, there were still discrepancies between the model and data. In particular, performance decreased in the model for the lowest rates, whereas it kept

increasing in the data. Also, performance was generally underestimated for the highest rates, especially for the Poisson distribution.

B. Interim discussion

A model of temporal pattern discrimination based on cross-correlation (Sorkin, 1990) was able to capture two important trends of the within-trial repetition-detection performance observed in our dataset: the better performance for long ICIs, and the better performance for Poisson-distributed compared to uniformly distributed ICIs. We interpret the model behaviour by considering the ratio between stimulus-related variance and internal-noise variance. Stimulus-related variance was beneficial to model performance, as it made patterns more distinguishable from each other for the cross-correlation operation. In contrast, internal noise reduced performance by corrupting the internal representation of the patterns. The important characteristic of the model here is that internal noise had constant variance. As interval durations decreased for high click rates, their representation in the model was proportionally more affected by internal noise, and the model performance decreased. As the intervals had higher stimulus-related variance for the Poisson distribution compared to the uniform distribution, their representation in the model was more robust to internal noise, and the model performance was higher for the Poisson distribution compared to the uniform distribution.

The main aim of the model was to gain an intuition about the effect of ICI distributions, so we kept its structure as simple as possible, but this implied some important limitations. In particular, this “toy” model had perfect memory. The only limiting factor for performance was the additive noise added to each interval, irrespective of interval duration or total number of intervals. An extension of the original model has been proposed including both additive and multiplicative noise (Sorkin, 1990). We tried including this multiplicative noise parameter, but it only provided an overall gain for the model predictions and did not change the qualitative differences between model and data. Another simplification is that the model assumed that participants could compare exactly the first and second halves of each stimulus, even though stimuli were presented continuously. As all trials had equal duration, it is not unlikely that participants learnt the approximate halfway position of the stimuli, and cross-correlation should be reasonably robust to imperfect estimates of the halfway point. Finally, we made no attempt at modelling the added benefit of repeated exposure for RefRC: this advantage was about

the same for all conditions, so this did not provide any useful constraint to add a parameter to the model.

The resulting fit was qualitatively correct, but still imperfect. The underestimation of performance at the lowest click rates could be due to issues with computing a meaningful correlation with a small number of intervals. The underestimation of performance at high click rates suggests that participants may have used other strategies (Ross & Houtsma, 1994) than the “holistic” comparison of all intervals implied by the correlation operation. For these high rates, it is plausible that some intervals, for instance the longest ones, were weighted more than others.

Finally, even if the model succeeded in predicting trends in the data, we do not wish to imply that the neural bases of click-train repetition detection literally involve computing cross-correlations over a broad range of time intervals. Auditory models based on auto-correlation could be used to detect repetitions using a similar strategy, but the long duration of our stimuli would place heavy demands on the required delays, and the applicability of such models to random click trains with purely temporal regularities has been challenged (Kaernbach & Demany, 1998). Rather, the present functional model served to identify a general limiting factor for the repetition-detection task, the balance between stimulus variability and internal noise. The underlying neural mechanisms remain as yet unspecified and some possibilities will be discussed in the following section.

2.7 Discussion

The effect of repeated exposure to random auditory temporal patterns was investigated in adult listeners, using an experimental paradigm that had previously been used to demonstrate learning for random spectro-temporal patterns (Agus et al., 2010). In three behavioural experiments, we showed that random auditory time patterns could be learnt rapidly, after only a few exposures. Moreover, the memory for temporal patterns was observed over a surprisingly broad range of inter-click intervals: from sparse click trains, containing about 5 clicks per seconds, to dense click trains, containing more than 50 clicks per second. Such a range encompasses the subjective transition between clicks being heard out individually, like in a rhythm, to clicks being fused together, like in a buzzing sound (Bendor & Wang, 2007; Pressnitzer et al., 2001). Immediate repetition

detection was easier for the rhythm-like patterns, but, surprisingly, this did not appear to affect the longer-term memory processes operating across trials, as the gain from increased exposure was comparable for all rates.

There were strong similarities between the characteristics of memory for noise (Agus et al., 2010) and clicks trains (present experiments). In both cases, learning could be achieved in an unsupervised and implicit manner, with the patterns to be learnt interspersed amongst other statistically matched sounds. The ancillary task used in our paradigm (within-trial repetition detection) was initially challenging. However, as the patterns were heard more than once over the course of a block, performance sometimes improved to reach almost-perfect performance in about 10 trials. This change in performance was interpreted as learning, as learnt patterns became perceptually easily distinguishable from novel ones (Agus & Pressnitzer, 2013).

There was, however, an important difference between the memory for noise and click trains. For noise, when a white noise exemplar was time-reversed after having been learnt, an almost-perfect transfer of memory was observed (Agus et al., 2010). Together with other converging evidence reviewed in the Introduction, this suggested that noise was learnt thanks to brief spectral cues. In contrast, here we observed that, when a time pattern was time-reversed after learning, there was no evidence of memory transfer. Rather, the time-reversed pattern had to be learnt afresh. Thus, for click trains, it was the extended series of time intervals and their precise ordering that was learnt.

This leads to an apparent puzzle: auditory memory for noise and click trains relied on qualitatively different cues, but still displayed similar characteristics. One parsimonious interpretation to resolve that puzzle is that both types of cues were recoded into a common neural code before being learnt. Let us first consider the case of noise. Noise contains spectral cues, which are encoded in the auditory system as spatial patterns of activity across frequency-tuned neurons: across auditory nerve fibres in the periphery, or across other neural populations in subsequent tonotopic maps at least up to primary auditory cortex (Bidelman et al., 2014). To form memory traces for such spatial patterns, standard Hebbian models of neural plasticity apply. Indeed, computational models have shown that neurons receiving inputs across a range of afferents, and equipped with well-established plasticity rules such as spike-time dependent plasticity, can develop selectivity to random patterns of near-simultaneously active afferents (Masquelier et al., 2008). If the afferents were frequency-selective neurons, then the scheme could apply to the learning of auditory noise (Viswanathan, 2016).

For auditory time patterns, the learning was dependent on details of successive time intervals and their order, and thus required the integration of non-simultaneous information. In the auditory periphery, at the level of the auditory nerve, time intervals are not encoded as a spatial map, but rather transmitted explicitly as inter-spike intervals (Cariani & Delgutte, 1996). Neural spikes synchronized to clicks have been observed up to at least primary auditory cortex for time-jittered click trains resembling our stimuli (Lu et al., 2001). However, there is also evidence that inter-spike intervals are recoded into a rate-place code: a “synchronised” neural population co-exists with “non-synchronised” populations, which encode click-rate through neural discharge rate in primary (Lu et al., 2001) and non-primary auditory cortex (Bendor & Wang, 2007) depending on the click-rates (Petkov & Bendor, 2016).

A rate-place code for click rate should be able to produce neural spatial maps distinguishing click trains with different rates. However, it has not been tested whether such a representation would support the learning of a specific pattern amongst other patterns with identical average rate, as was the case in our experiments. To do so, a generic class of computational models have been put forward: state-dependent networks, which, thanks to the complex temporal dynamics imposed by individual neurons and by the network topography, can produce a large number of unique network states depending on the exact temporal succession of input spikes (Lim, Lagoy, Shinn-Cunningham, & Gardner, 2016). Intuitively, such a network state is simply a place pattern of neural activity that depends on the past and present input, with the potential to distinguish between temporal patterns even when they share the same average rate. Such a model has been used to account for behavioural context effects in auditory temporal pattern perception in human listeners (Karmarkar & Buonomano, 2007). Moreover, recent physiological and behavioural evidence for a conversion from time patterns to place patterns have been found in the avian auditory system. Using random click trains with time intervals between 11 ms and 40 ms—intended to mimic the timing characteristics of bird sounds, but not unlike our own stimuli—Lim et al. (2016) recorded neural activity in primary and secondary auditory regions of the anaesthetized zebra finch forebrain. They observed distinctive neural place patterns of activity in secondary regions only, which could support the behavioural discrimination between acoustic temporal patterns. They also tested for memory transfer after time-reversal of the click trains: there was no transfer in a behavioural task, consistent with observed changes in neural place maps after time-reversal.

In summary, such a general scheme would in principle be consistent with all aspects of the present results. Somewhere along the auditory pathways, time patterns could be converted to neural place patterns, unique to each input sequence and sensitive to time-reversal. Once recoded as place patterns, unsupervised learning of acoustic temporal patterns could be achieved through standard neural plasticity processes (Masquelier et al., 2008). As the spectral cues present in noise or tone clouds (Agus et al., 2010) could also be coded as neural place patterns, the same plasticity processes apply, hence explaining the similarities observed for the learning of all types of auditory stimuli. There are of course many aspects of our findings that would need to be tested more thoroughly against this speculative account. In particular, we observed learning over a broad range of time scales, straddling the divide between “synchronized” and “non-synchronized” neural populations (Lu et al., 2001). The time scales involved in the state-dependent models (Lim, Lagoy, Shinn-Cunningham, & Gardner, 2016) are broadly consistent with the stimulus range we tested, but differences between stimuli and species prevent a more detailed comparison. The locus, or loci, of this putative recoding and subsequent learning remain to be ascertained in humans. Interestingly, the available fMRI (Kumar et al., 2014) and EEG (Andrillon et al., 2015) data for the learning of noise suggest neural correlates in secondary auditory areas, which would be consistent with the animal data (Lim, Lagoy, Shinn-Cunningham, & Gardner, 2016; Lu et al., 2001).

2.8 Conclusions

Using random click trains in an auditory memory paradigm (Agus et al., 2010), we showed that auditory temporal patterns could be learnt by adult human listeners. We tested average click rates from about 5 clicks per seconds up to about 50 clicks per second. Even though this range covered patterns sounding like rhythms for low rates or irregular buzzing sounds for high rates, learning was observed in all cases. We also varied the higher-order statistics of the random distribution of time intervals. A higher variance led to better within-trial repetition detection, but did not affect much the across-trial longer term memory effect. In all cases, learning was fast and occurred in an unsupervised fashion, with the patterns to be learnt interspersed among statistically matched patterns. The learning depended on the precise time-series of interval and their presentation order, as time-reversal disrupted performance. We suggest that auditory

temporal patterns may be recoded as neural spatial patterns along the auditory hierarchy, which are then amenable to standard neural plasticity mechanisms. Overall, the experimental evidence now shows that audition benefits from remarkably fast and efficient unsupervised learning processes for noises (Agus et al., 2010), other random spectro-temporal patterns like tone clouds (Kumar et al., 2014), and also purely temporal patterns such as click trains (present data). This is likely to cover any sort of acoustic cues present in natural sounds, e.g., the timbre of a novel sound source (Pressnitzer et al., 2015). More generally, such an almost-inevitable emergence of memory traces after repeated exposure, for different kinds of auditory cues and several time scales, would be consistent with views suggesting that memory may be viewed as an integral part of sensory processing (Hasson et al., 2015).

Acknowledgments

We would like to thank Christian Kaernbach and an anonymous reviewer for helpful comments on a previous version of the manuscript. This work was funded by the ERC Grant ‘ADAM’ No. 295603, the ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*.

Chapter 2 Appendix

In all experiments, we distinguished good blocks, where learning of RefRC was observed, from bad blocks, where no learning of RefRC was observed. Here, we further analyze the temporal patterns that comprised the RefRCs, to test whether there were systematic acoustic differences between the two kinds of blocks.

Even though generative parameters were the same for all blocks, it is possible that by chance good blocks were associated with fewer time intervals to learn, or perhaps more distinctive ones. For each stimulus used in the experiments, we computed the mean ICI, M , inversely related to the number of intervals in the pattern as duration was kept constant. We also computed the standard deviation of ICIs, SD . In the main results, higher M s and SD s were associated with better performance. Finally, we computed the maximum of the absolute value of z -scored ICIs (using an estimation of standard deviation of all RefRC intervals within a condition). This later index should be sensitive to exceptionally long or exceptionally short ICIs within a pattern, which could provide distinctive cues.

All indices were then averaged within rate conditions for Experiments 1 and 2. To assess the reliability of the differences between good and bad blocks, we used Welch's two sample t -tests, as samples had unequal sizes. We report both exploratory, non-corrected p -values with a significance criterion set at $p < 0.05$, and p -values corrected for multiple comparisons using the Bonferroni method, with the same significance criterion.

Table 1. Observed characteristics of RefRCs corresponding to good and bad blocks. For each RefRC pattern, three indices were computed: M , the average ICI; SD , the standard deviation of ICIs; $\text{Max}(|z|)$, the maximum absolute value of z -scored ICIs. Indices were then averaged for each rate condition and experiment. Values in bold and italics indicate good/bad contrasts reaching a statistical significance criterion (Welch's t -test, uncorrected. Bold: $p < 0.01$. Italics: $p < 0.05$)

Rate	5		10		18		33		57	
	Good	Bad	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Block group	36	20	31	25	23	33	22	34	13	43
Number of blocks										
M	<i>0.218</i>	<i>0.255</i>	<i>0.101</i>	<i>0.114</i>	<i>0.056</i>	<i>0.056</i>	<i>0.029</i>	<i>0.031</i>	<i>0.018</i>	<i>0.018</i>
SD	0.101	0.088	0.055	0.054	0.025	0.025	0.012	0.012	0.005	0.004
$\text{Max}(z)$	1.097	0.872	1.540	1.430	1.759	1.723	1.763	1.759	1.735	1.747
Number of blocks	44	12	51	5	35	21	32	24	21	35
M	0.182	0.285	0.110	0.136	0.058	0.057	0.030	0.031	0.018	0.018
SD	0.194	0.147	0.089	0.124	0.046	0.041	0.021	0.019	0.008	0.008
$\text{Max}(z)$	1.255	1.406	<i>1.745</i>	<i>2.574</i>	2.737	2.276	3.169	2.672	3.758	3.529

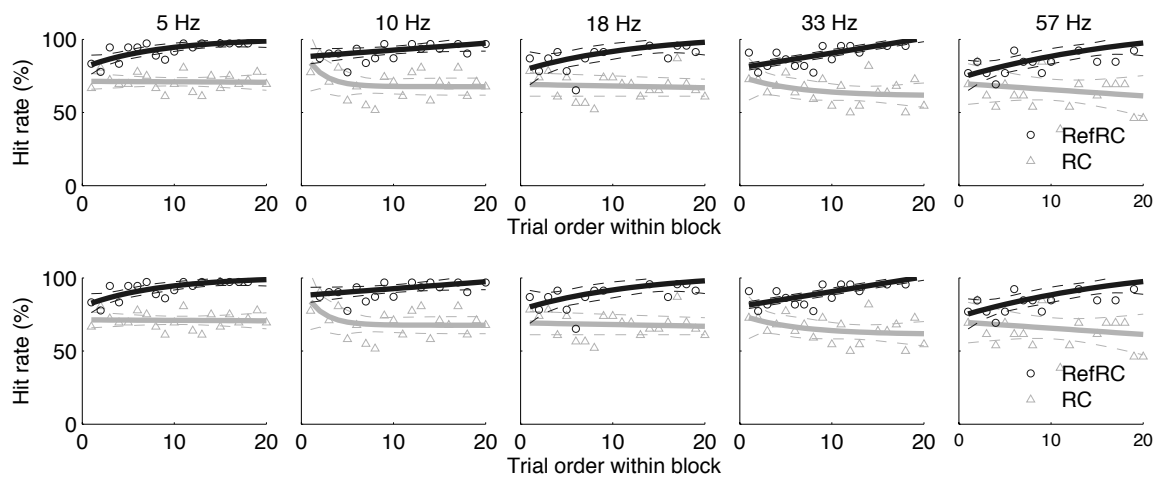
Results are shown in Table I. Most comparisons did not reach our lenient uncorrected significance criterion. For Experiment 1, only the contrast between mean ICIs at rates 5 Hz ($t(39.64) = 2.2, p < 0.05$) and 10 Hz ($t(41.67) = 2.28, p < 0.05$) reached significance. For Experiment 2, only the contrasts between mean ICIs at 5 Hz ($t(14.47) = 3.93, p < 0.01$) and between Max(|z|) at 10 Hz ($t(6.85) = 3.48, p < 0.05$) and 33 Hz ($t(51.03) = -2.04, p < 0.05$) reached significance. When applying the correction for multiple comparisons, only the contrast between mean ICIs at 5 Hz in Experiment 2 remained significant ($p < 0.05/15$).

The only significant difference that survived the multiple comparisons correction was found for low click rates. This may have been expected, as low rates correspond to small numbers of intervals, and hence a greater likelihood to obtain discrepancies between generative and observed characteristics for a given random draw. In this case, good blocks were actually associated with higher observed rates. This is in apparent contradiction with the general pattern of results, where lower rates led to better performance. It is possible that for the lowest rate here, the limiting factor was that patterns became too sparse to be distinguishable.

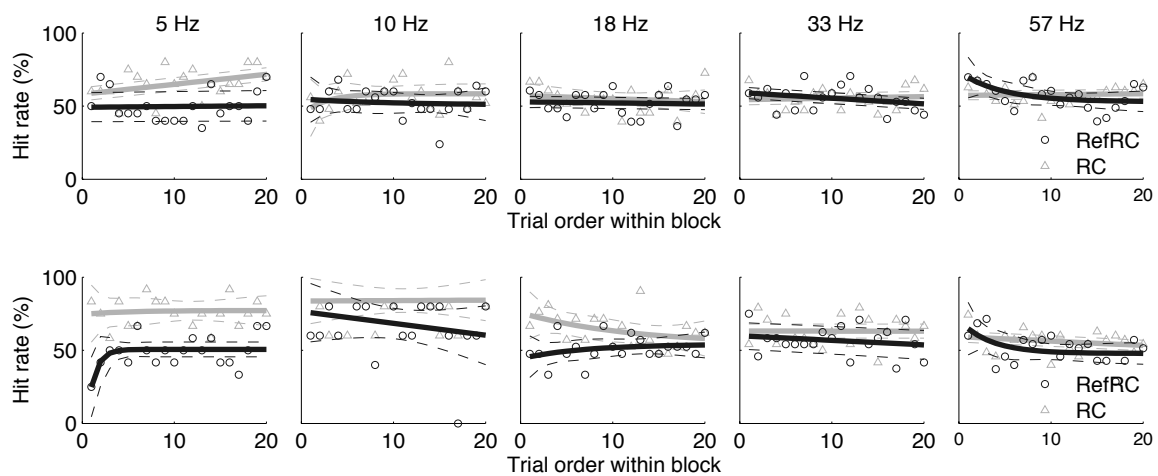
In any case, we interpret Table 1 as suggesting that, over all experimental conditions, good and bad blocks were not systematically different in terms of mean ICIs, standard deviation of ICIs, and unusually short or long ICIs.

AUDITORY MEMORY FOR RANDOM TIME PATTERNS

Supplementary information



2S-1. Learning of good blocks across all rates for Experiment 1 (top) and Experiment 2 (bottom). The solid lines are best-fit exponentials for each stimulus condition (grey triangles: RC, black circles: RefRC), with dashed lines indicating 95% confidence interval.



2S-2. Learning of bad blocks across all rates for Experiment 1 (top) and Experiment 2 (bottom). The solid lines are best-fit exponentials for each stimulus condition (grey triangles: RC, black circles: RefRC), with dashed lines indicating 95% confidence interval.

Chapter 3

Memory for random time patterns in audition, touch, and vision

This chapter has been submitted as:

Kang, H., Lancelin, D., & Pressnitzer, D. (submitted). Memory for random time patterns in audition, touch, and vision. *Neuroscience*.

Parts of this work have been presented as:

Kang, H., Lancelin, D., & Pressnitzer, D. (2017). A Memory Process of Temporal Information in Sensory Modalities: Audition, Vision, and Touch. 40th MidWinter Meeting of the Association for Research in Otolaryngology, Baltimore, USA (*Poster*).

Kang, H., & Pressnitzer, D. (2016). Memory of Temporal Patterns in Different Sensory Modalities: Audition, Vision, and Touch. Basic Auditory Science Meeting, Cambridge, UK (*Poster*).

3.1 Abstract

Perception deals with temporal sequences of events, like series of phonemes for audition, dynamic changes in pressure for touch textures, or moving objects for vision. Memory processes are thus needed to make sense of the temporal patterning of sensory information. Recently, we have shown that auditory temporal patterns could be learnt rapidly and incidentally with repeated exposure [Kang, Agus, & Pressnitzer, J. Acoust Soc. Am. 2017]. Here, we tested whether rapid incidental learning of temporal patterns was specific to audition, or if it was a more general property of sensory systems. We used a same behavioral task in three modalities: audition, touch, and vision, for stimuli having identical temporal statistics. Participants were presented with sequences of acoustic pulses for audition, motion pulses to the fingertips for touch, or light pulses for vision.

MEMORY FOR RANDOM TIME PATTERNS IN AUDITION, TOUCH, AND VISION

Pulses were randomly and irregularly spaced, with all inter-pulse intervals in the sub-second range and all constrained to be longer than the temporal acuity in any modality. This led to pulse sequences with an average inter-pulse interval of 166 ms, a minimum inter-pulse interval of 60 ms, and a total duration of 1.2 s. Results showed that, if a random temporal pattern re-occurred at random times during an experimental block, it was rapidly learnt, whatever the sensory modality. Moreover, patterns first learnt in the auditory modality displayed transfer of learning to either touch or vision. This suggests that sensory systems may be exquisitely tuned to incidentally learn re-occurring temporal patterns, with possible cross-talk between the senses.

3.2 Introduction

Audition, touch, and vision sense different parts of the physical world, such as acoustic pressure waves, surface textures, or light patterns. Peripheral sensory receptors in each modality are adapted to best transduce their respective physical input (e.g. Lewicki, 2002), and, from subsequent neural processes, qualitatively distinct perceptual qualities emerge, such as auditory timbre, tactile texture, or visual color. However, at a more general level of description, all sensory modalities are embedded in time: they all must deal with sequences that contain possibly crucial information in their temporal patterning (O'Regan, 2011). Thus, memory processes applying to temporal sequences would seem beneficial in any sensory modality, and may also have to address similar computational constraints across modalities (Hardy & Buonomano, 2016).

Recently, we documented a rapid form of incidental learning for temporal sequences in the auditory modality (Kang et al., 2017). Adapting a paradigm previously used in audition (Agus et al., 2010) or vision (Gold et al., 2014), we observed fast incidental learning of time patterns made of irregularly spaced audio clicks, with inter-click intervals in the range of tens to hundreds of milliseconds. Learning occurred, in an unsupervised manner, as long as these patterns re-occurred over the course of an experiment. In this recent study, we ensured that the only cues available to learn patterns were in the precise sequence of time interval durations, and not in any other auditory-specific cue. So, the same experimental paradigm may be transposed to other sensory modalities, simply by conveying time intervals not through audio clicks, but rather through modality-adapted events. This is what was done in the present study. Our aim was to investigate whether rapid incidental learning of random time patterns was specific to audition, or if it could be a more general feature of perceptual systems.

On a neural level, for stimuli consisting of time intervals delimited by brief energy pulses, the input time patterns will be reflected in neural spike time patterns in peripheral receptors and also at higher stages of processing, at least for moderate rates of up to a few pulses per second. Temporal patterning of spike trains has been observed in sensory cortices for audition (Lu et al., 2001), touch (Saal et al., 2016), and vision (Gur & Snodderly, 1997). Of course, there are differences in temporal resolution between these modalities. Also, it is a matter of controversy whether spike time patterning is an epiphenomenon of peripheral encoding (Gur & Snodderly, 1997; Salinas et al., 2000) or a

true information-bearing code (Cariani, 2001; VanRullen et al., 2005). But at a minimum, the possibility exists to learn temporal patterns of spikes at the cortical level in all of the three modalities considered here, audition, touch, and vision. Probing the behavioral learning of purely temporal patterns in these modalities would be a first step in investigating the underlying neural mechanisms.

An obvious question then arises: if learning of time patterns were to be observed in all modalities, would it be the result of canonical computational principles, or rather reflect mechanisms specific to each modality of entry? The question overlaps with broader issues in time research, such as whether the psychological and neural representations of time are modal or amodal (Ivry & Schlerf, 2008; van Wassenhove, 2009; Grondin, 2014). Several studies have approached the issue by inducing perceptual learning for a time interval in one modality, and then test for transfer of learning in another modality. A variety of results have been found, often with asymmetric transfer characteristics across modalities (e.g. Nagarajan et al., 1998; Lapid et al., 2009; Bratzke et al., 2012; Pasinski et al., 2015). As summarised by Pasinski et al. (2015), these results may reflect differences in task demands. Using a variant of the time-interval discrimination task, they found a behavioral advantage of the auditory modality over the visual modality, but similar expectancy-related and memory-related EEG responses across the two modalities (Pasinski et al., 2015), consistent with a combination of modality-specific mechanisms for the encoding of time intervals followed by modality-general memory processes (see also Merchant et al., 2008).

It is yet unclear whether such a conclusion would hold beyond interval discrimination tasks, for instance for the learning of more complex temporal patterns. There are behavioral demonstrations of rapid learning for complex sequences in audition, touch, and vision separately (Gold et al., 2014; e.g. Bale et al., 2017; Kang et al., 2017), but not all stimuli in these studies were purely temporal patterns. Sequence learning has been compared across modalities (Handel & Buffardi, 1968; Manning et al., 1975; Conway & Christiansen, 2006), but again combining timing-cues with a variety of modality-specific cues such as sound frequency or spatial location, which may impact results in unsuspected ways (e.g. Grahn, 2012). Perceptual learning of purely-temporal rhythmic sequences has been compared between audition and vision (Patel et al., 2005; Grahn, 2012; Barakat et al., 2015), but rhythmic sequences may recruit additional, beat-based mechanism for sequence learning (e.g. Pasinski et al., 2015). In one example using comparable stimuli across modalities, which were aperiodic and differing only by timing

characteristic, Nazzarro & Nazzarro (1970) had participants learn auditory or visual Morse code “words”. They found faster learning in the auditory modality. This finding echoed early suggestions of an advantage of audition over vision for the motor reproduction of temporal rhythms (Fraisse, 1948), and also classic findings of a greater temporal acuity for audition compared to touch or vision for single-interval discrimination (Goodfellow, 1934). Interestingly however, when the discriminability between elements was equated for auditory and visual Morse, the auditory advantage vanished (Hansen & Cottrell, 2013). Note also that the Morse code task involved explicit learning, which could recruit different mechanisms than incidental learning (e.g. Chen & Zhou, 2014). In summary, while there is ample evidence for behavioral sequence learning in audition, vision, and touch, less is known about the learning of purely temporal complex patterns. In particular, a detailed comparison of the learning of aperiodic temporal patterns in all three modalities, with the same participants and procedure, is lacking.

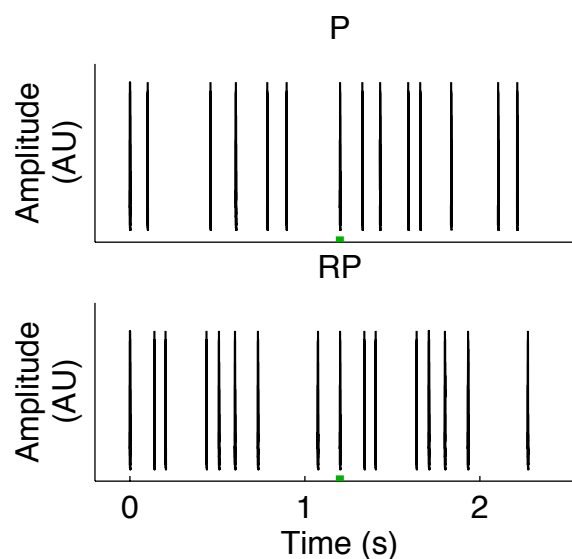


Figure 3-1. Schematic illustration of the sequences used in Experiments 1 and 2. For random pulse sequences (P), pulses delineated time intervals drawn from a Poisson distribution with a 7 Hz rate and a 60 ms refractory period, for a total sequence duration of 2.4 s. For the repeated pulse sequence (RP), a 1.2 s sequence was generated as for P, but seamlessly repeated to obtain a 2.4 s sequence. The green square indicates the midpoint of the sequences. Participants had to discriminate P from RP. In addition, a third stimulus type was included (Reference repeated pulse sequence, RefRP), generated as RP but re-occurring over several trials at random times during an experimental block. Pulses were audio clicks in audition, tangential motion pulses to the fingertips for touch, and light flashes for vision.

In the present set of experiments, we adapted the auditory paradigm of Kang et al. (2017) to the tactile and visual domain. Briefly, we used random temporal patterns made of irregular time intervals delineated by brief energy bursts. The energy bursts were audio clicks for audition, tangential motion pulses applied to the fingertips for touch, and light flashes for vision. The task as explained to participants was to discriminate two types of stimuli: either a fully random pulse sequence lasting 2.4 s (pulses, P), or a 2.4 s pulse sequence made from the seamless repeat of two identical 1.2 s sequences (repeated pulses, RP). The two conditions are illustrated in Figure 3-1. In addition, without instructing participants, we also introduced a third type of stimulus: reference repeated pulses (RefRP), which were constructed exactly as RPs but re-occurred identically over randomly selected trials throughout the course of an experimental block. Thus, participants were exposed to the exact same inter-pulse interval (IPI) pattern for several trials for RefRP, whereas the IPI patterns comprising P or RP stimuli were unique to one trial. An advantage in performance for RefRP over RP, that is, for re-occurring patterns over novel patterns, can be interpreted as perceptual learning (Agus et al., 2010; Agus and Pressnitzer, 2013; Luo et al., 2013; Gold et al., 2014; Andrillon et al., 2015; Kang et al., 2017).

Importantly, here we ensured that the encoding of temporal patterns was not limited by temporal acuity in any of the three modalities tested (Goodfellow, 1934). A pre-test served to adjust the minimum IPI required for all modalities. The IPI distribution statistics was then fixed across modalities. A similar approach had been taken by Marks (1987), who collected perceptual similarity judgments for temporal sequences in audition, touch, and vision, with identical supra-thresholds IPIs. He concluded that, in such a setting, the perceptual dimensions underling similarity judgments were common to all three modalities (Marks, 1987). Here, we followed the same logic to probe perceptual learning rather than perceptual similarity. Our stimuli had identical statistical properties in all modalities and performance was compared across modalities using the same task and participants.

We also tested for transfer of learning across modalities, to further address the issue of modality-specific or modality-general mechanisms of perceptual learning. Participants performed the same task as for the first experiment, but switched modality of presentation of the time patterns mid-block. In contrast to previous transfer studies, here participants were tested on incidental and not explicit learning, using complex and aperiodic sequences. We limited the modality pairings to initial learning in audition, before transfer

to either touch or vision, as audition provided the most robust learning effects in the first experiment. The experimental hypothesis was that, if participants retained a high performance on learnt RefRP patterns just after the modality switch, this could indicate transfer of learning.

3.3 Experimental procedures

3.3.1 General information common to all experiments

Participants

A total of 53 volunteers took part in the Pre-test, Experiment 1, and Experiment 2. They were typical adult participants (age in years $M = 25.0$, $s.d. = 2.8$). All participants had normal hearing, with measured pure-tone thresholds lower than 20 dB HL at frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. All participants reported normal or corrected to normal vision and no history of mental disorder or epilepsy. None had previous experience with similar experiments and they were all naïve to the goal of the study. Participants gave informed written consent on their participation and they were compensated 10 euros per hour at the end of the experiment. The study was approved by the ethical committee of U. Paris Descartes, France (CERES, IRB: 20154000001072).

Apparatus

The stimuli were generated in Matlab 2014b (Mathworks) on a MacBook Pro. They were generated at a 44.1 kHz sampling rate and 16 bit resolution, then converted to analog signals with an RME Fireface UC soundcard.

For the Audition conditions, the analog signals were played as sounds over Sennheiser HD 215 headphones, diotically.

For the Touch conditions, the same signal generation method was applied, but the output of the soundcard was split into two channels. The main stimulus channel was sent to a custom-made tactile stimulation box. The box contained a vibrotactile transducer (TL-002-14R, Haptuator Redesign, Tactile Labs) attached to a mobile 3D-printed plastic plate, resting on ball bearings covered by elastic bands. Inside the box was a custom amplifier (Tactile Labs Haptu amp) driving the Haptuator. The overall box dimensions were 12 cm width by 12 cm height by 3.5 cm depth. It was fixed on a mouse pad with a

gel wrist rest. Participants could comfortably rest their hand and put their fingertips on the plate, which moved tangentially to their fingertips when the Haptuator was driven. The second audio channel was used to play noise through headphones, in order to mask any sound produced by the tactile stimulation box (see Pre-test, Stimuli).

For the Vision conditions, the output of the sound card was again split into two channels and the main stimulus channel sent to a custom-made visual presentation device. The device consisted of a 3D-printed opaque box, of dimensions 11.3 cm width by 11.3 cm height by 5.3 cm depth, with a circular central aperture of diameter 5 cm in the center. The aperture was covered by a translucent panel. Behind the panel, 4 green LEDs (Osram LT W5AM) were positioned in a cross pattern, each at a distance of 1.27 cm to the aperture center. This provided homogenous lighting of the translucent panel. With the experimental configuration, the panel covered about 5° of visual angle in the center of the participants visual field. The output of the soundcard was amplified through a custom-made circuit to control the LEDs in a binary fashion (on or off) with high timing accuracy. The second channel of the audio card was again used to play the masking noise over headphones, for consistency across conditions.

All experiments took place in a soundproof booth (IAC 400-A series, single-walled) with moderate lighting (10 cd/m²) provided by DC-lamps to avoid visual flicker due to the mains. Participants were seated at 57 cm of the screen or visual stimulation box, depending on the condition. They used a chin-rest for all conditions and provided their responses by mean of a computer keyboard located to their right. During the Touch conditions, they rested their left hand on the tactile box. Participants wore headphones in all conditions.

Stimuli

Stimuli consisted of irregular pulse trains with random IPIs. The IPIs were drawn from a Poisson distribution with refractory period, with the parameters of the distribution adjusted during the pre-tests and subsequently kept fixed for all modalities and for Experiments 1 and 2.

For Audition, the pulses were audio clicks of duration 0.2 ms. Pulse trains were high-pass filtered at 2 kHz to prevent the appearance of spectral cues after peripheral auditory filtering (Kang et al., 2017). Pink noise was added to the filtered click trains, to mask any potential distortion product but also for consistency with other conditions (see Pre-test, Results). The pink noise was generated in the frequency domain between 50 Hz

and 2 kHz. It started 0.2 s before the pulse sequence stimulus and ended 0.2 s after the pulse sequence.

For Touch, the pulses were tangential motion pulses driven by rectangular voltage steps of duration 0.2 ms. To make the best use of the dynamic range of the soundcard with respect to the characteristics of the Haptuator, the driving step trains were low-pass filtered at 400 Hz. Audio pink noise was presented over headphones, with the same level and characteristics as in the Auditory condition, to mask the sound produced by the tactile stimulation box.

For Vision, the pulses were light flashes driven by rectangular voltage steps of duration 0.2 ms. Here, the driving steps were unfiltered. Audio noise with the same characteristics as in the other two conditions was simultaneously presented over headphones.

3.3.2 Pre-test part I: Setting the masking noise level

Participants

There were $N = 6$ participants (2 male, age in years $M = 26.0$, $SD = 3.08$).

Stimuli

The level of acoustic noise needed to mask the sound produced by the tactile box was measured. A tactile pulse train, presented over the tactile box, was presented simultaneously with an audio pink noise, presented over headphones. Based on pilot data and several iterations of the pre-tests, a rate parameter of 7 Hz and refractory period of 60 ms was selected for the Poisson distribution of IPIs of the tactile pulse train. This resulted in an average IPI rate of 6 Hz. A signal-to-noise ratio (SNR) was defined as the ratio of the root-mean square (*rms*) value of the signal driving the tactile box, which remained fixed, over the *rms* value of the signal driving the headphones, which was varied. The SNR tested were [-2, 0, 4, 12, 28] dB, which corresponded to a level of the acoustic noise of [56, 54, 50, 46, 42, 29] dB SPL, A-weighted.

Procedure

Participants were sat in the sound booth, in the experimental setting for all experiments with their head on the chinrest and headphones on. The experimenter sat next to the participant and put one hand on the tactile box, to reproduce the pressure of

the hand of the participant on the box during the actual experiments. For each SNR, 40 trials were presented (20 audio only, 20 audio and tactile box), for a total of 200 trials presented in random order. The participant's task was to respond whether a tactile stimulus was played or not, a task which could only be achieved with auditory detection of the tactile stimulus.

3.3.3 Pre-test part II: Setting the minimum IPI

Participants

There were $N = 5$ participants (2 male, age $M = 26.7$, $SD = 2.50$). One participant was re-invited from the Pre-test part I, the others were new participants.

Stimuli

The minimum IPI of pulse trains was varied for touch and vision, in order to measure the shortest IPI for which time intervals could be perceived in all modalities. It is known that audition has a much better temporal resolution than vision or touch (e.g. Gescheider, 1966) so this modality was not included in the pre-test. Each trial consisted of a target stimulus, containing two pulses, and a comparison stimulus containing one pulse. The target stimuli consisted of two successive 0.2 ms pulses, separated by an IPI that was varied across trials. The IPIs tested were [25, 27, 31, 39, 55, 87] ms. The comparison stimulus consisted of a single 0.4 ms pulse, which correspond to twice the duration of a single target pulse. Target and comparison stimuli were separated by a time gap of 500 ms and presented in random order. Masking noise with the level set in the pre-test part I was simultaneously presented over headphones.

Procedure

Participants were sat in the sound booth, using the chinrest. For touch, their hand rested on the tactile box. For vision, they fixated the visual stimulation box. Their task was to report which presentation interval contained the target, two-pulses stimulus. There were 26 trials per IPI condition, for a total of 156 trials, presented in random order. Blocks for each modality were run separately in random order.

3.3.4 Experiment 1

Participants

For Experiment 1, there were $N = 18$ participants (7 male, age $M = 25.2$, $SD = 3.4$). All participants were tested on all modalities.

Stimuli

For Audition, the pulse trains plus masking pink noise were presented over headphones as described above. Based on the results of the pre-tests, we set the rate parameter of the Poisson distribution of IPIs at 7 Hz, with a refractory period of 60 ms, corresponding to a nominal rate of 6 Hz. Stimuli (clicks plus noise) were presented at an average level of 65 dB SPL, A-weighted, as calibrated with a Bruel & Kjaer Type 2250 sound-level meter set on “slow integration” connected to a Bruel & Kjaer Type 4153 artificial ear.

For Touch, pulse trains were presented to the fingertips through the vibrotactile box, with the same IPI parameters as for the audition condition. This resulted in a maximum measured acceleration value per pulse of 0.045g on average (measured with a GY-61 accelerometer connected to a Nucleo STM32 acquisition board). Masking noise was delivered through headphones at a level of 56 dB SPL, consistent with the level set with the Pre-test part I and identical to the noise used in the Audition condition.

For Vision, pulse trains were presented in the center of the visual field, again with the same IPI parameters as for the Audition and Touch conditions. The luminance of a single visual pulse was on average of 16 cd/m^2 , as calibrated with a Konica Minolta C-100A Luminance and Color meter. The same masking noise as in the two preceding conditions was delivered through headphones.

Procedure

The procedure was the same in all modalities. It replicates a previous study in the auditory modality (Kang et al., 2017). A qualitative description of the task is also provided in the Introduction and only specific details are provided here.

There were 80 trials per block (40 P trials, 20 RP trials, 20 RefRP trials) presented in pseudo-random order, with the only constraint that no consecutive trials should present RefRPs. No feedback was given. For each modality, four blocks were run, corresponding to 4 different RefRPs (different across listeners). Each modality was run in successive blocks, with an order counterbalanced across participants. The whole experiment was

completed in two 2-hours sessions, over successive days.

Training and inclusion criterion

For each modality, a training session was provided before data collection began (Agus et al., 2014) to familiarize participants with the within-trial repetition detection task. A first training block used sequences containing 8 seamless repetitions of a 1.2 s pulse train, to be discriminated from sequences of the same overall duration but containing no repeat. This was intended as an easy condition, to illustrate the repetition detection task. Only 10 trials were run for this easy condition. Then, the number of within-trial repetitions was reduced for further training blocks, first to 4 repeats, then to 2 repeats. Each of these training blocks had 40 trials (20 RP, 20 P). Note that the RefRP stimulus type was never introduced during training. Auditory feedback on response correctness was provided after each trial.

Participants who scored below chance in any of these blocks were told that the experiment would be terminated. This only happened for 2 participants, both in visual blocks.

3.3.5 Experiment 2

Participants

There were $N = 24$ participants (6 male, age $M = 24.5$, $SD = 3.40$). They were randomly assigned to one of two experimental conditions, auditory-touch or auditory-vision.

Stimuli

Experiment 2 examined transfer of learning, so each block consisted of a first half in one modality (Audition) and a second half in another modality (Touch or Vision depending on the experimental group). The stimuli within each modality were identical to Experiment 1.

Procedure

Each block started identically to the audition condition of Experiment 1, with 80 trials (40 P, 20 RP, 20 RefRP). At the end of the 80 trials, an auditory beep signaled to participants that they had to switch modality. They pressed a keyboard key to indicate their readiness, and presentation continued in the assigned modality (touch or vision). In four blocks, the

RefRP of the second modality had identical IPIs as the RefRP in the first auditory modality (Transfer condition). In four different blocks, the RefRP of the second modality had different IPIs from the RefRP of the first auditory modality (Control condition). Transfer and Control blocks were run in random order.

A training session on each modality (either Audition and Vision, or Audition and Touch) was provided before data collection began, as in Experiment 1. All participants met the inclusion criterion. The experiment took about 2 hours and was run in a single session.

3.3.6 Statistical analyses

Pre-test

For the Pre-test part I, which was a yes/no task, we computed the sensitivity index d' of signal detection theory (Macmillan & Creelman, 2004) for each block. The index d' was then averaged for each participant and SNR value, and the resulting distributions over participants compared to the chance value of 0 using t -tests. For part II, which was a 2AFC task, we computed the percentage correct for each IPI condition, averaged the value for each participant, and compared the resulting distributions with the chance value of 50% using t -tests.

Experiment 1

Performance was summarized using d' (Macmillan & Creelman, 2004) for each block and stimulus type (RP or RefRP). To do so, hits were defined as “repeated” responses for RP or RefRP, and false alarms were defined as “repeated” responses for P. Then, d' was averaged over all blocks for each participant, stimulus type, and modality. A repeated-measures analysis of variance (ANOVA) was performed on the dependent variable d' , with within-subject factors of stimulus type (RP or RefRP) and modality (Audition, Vision, or Touch), and with random factor of participants. Mauchly's test for sphericity confirmed that the data met the assumptions required for ANOVA (all $p > 0.05$). The level of statistical significance was set at $p < 0.05$ and the generalized eta-squared measure η^2_g of effect size was also computed. Post-hoc paired t -tests were performed to compare the RP and RefRP conditions within each modality. Holm-Bonferroni correction was applied for multiple comparisons when appropriate (specified in the text).

The time course performance on different conditions was also evaluated, to test

whether any advantage of RefRP over RP could be attributed to improved performance with exposure, or to a chance selection of particularly easy RefRP exemplars from the outset. To do so, we adapted the method described in Agus et al. (2010) and Kang et al. (2017). Responses were converted to hit rate for RP and RefRP and aggregated in trial bins over all blocks and participants, for each modality. The first bin represents the first time participants heard e.g. a RefRP during a block. These average values were fitted either with a flat line corresponding to the average performance (1-parameter model) or an exponential function (3-parameters model) using a least-squares method. The goodness of fit was estimated by the least-squared error and a penalization for extra parameters (Motulsky & Christopoulos, 2004). If the exponential model was selected, we concluded that performance evolved over time.

Experiment 2

The analysis of Experiment 2 focused on performance just after the switch across modalities, to evaluate performance in the new modality before any new learning could have taken place. The time course of performance was summarized as in Experiment 1, by computing hit rate for the different stimulus types according to their trial order. We then computed the average hit rate for the first 5 trials for RefRP in the second modality (either Touch or Vision), just after the switch. In the Transfer condition, the IPIs comprising the RefRP had been previously heard in Audition. An advantage of RefRP for the Transfer over the Control condition for those initial first 5 trials was thus taken as a measure of transfer of learning. A mixed-design ANOVA Type III (because the number of subjects became unbalanced after selecting blocks with initial learning before the modality switch) was performed on the RefRP hit rate, with a within-subject factor consisting of condition (Transfer or Control), a between-subject factor consisting of modality (Touch or Vision), and a random factor consisting of participant. Post-hoc paired *t*-tests were further applied to compare performance between Transfer and Control conditions. Finally, a similar ANOVA was run for the first 5 trials of the first part of the block, in Audition, to check that the random distribution of RefRPs across Transfer and Control condition produced balanced results in audition.

3.4 Results

3.4.1 Pre-test

The pre-tests were designed to determine the level of auditory masking noise required to mask any sound produced by the tactile stimulation box (part I) and to measure the minimum IPI ensuring that all time intervals in the patterns would be above the perceptual fusion threshold in all modalities (part II).

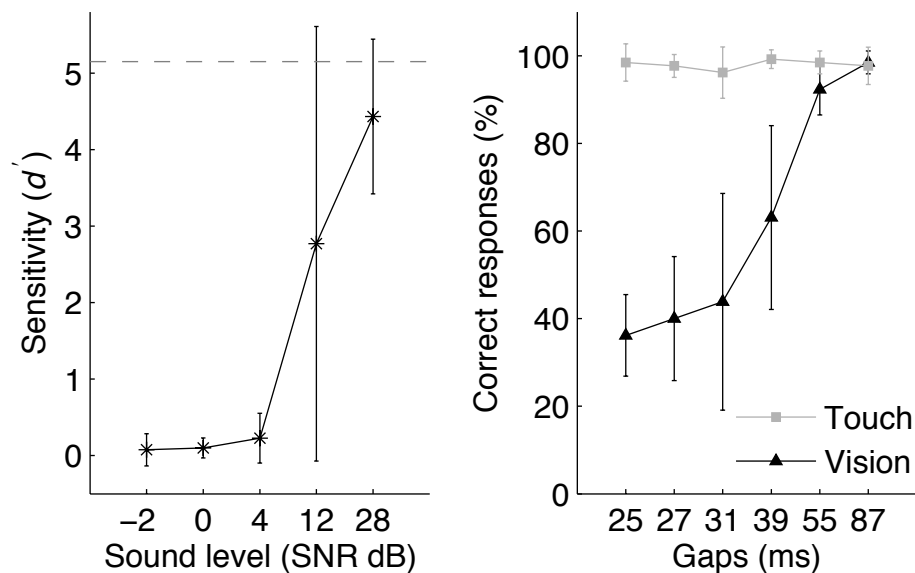


Figure 3-2. Results for the Pre-Test. *Left*: The auditory detectability of tactile stimuli, in d' units, is shown as a function of the signal-to-noise ratio between the signal driving the tactile stimulation box (fixed) and the signal driving headphones to play simultaneous auditory masking noise (variable). The main experiments were run with a SNR of -2 dB, corresponding to a masking noise at 56 dB SPL, A-weighted. *Right*: The percent correct for a one versus two pulses discrimination task is shown as a function of the time gap between pulses. The main experiments were run with a minimum inter-pulse interval of 60 ms.

For part I, we measured the ability of participants to detect the sound produced by the presentation of the tactile stimulus, in presence of varying levels of simultaneously-presented acoustic noise. Figure 3-2, left, illustrates detection performance a function of SNR. The SNR was defined as the ratio of the *rms* value of the electrical signal driving the tactile stimulator, over the *rms* value of the electrical signal driving the headphones.

As expected, the auditory detection of the tactile stimulus became more difficult with increasing levels of auditory noise, that is, for decreasing SNRs. Performance was almost perfect for high SNRs, then decreased with notable variability across participants for medium SNRs, and then consistently converged toward chance for low SNRs. At the lowest SNR, the mean d' value was 0.076, which was not different from the chance value of 0 (one-sided t -test, $t(4) = 1$, $p = 0.187$). We conservatively selected this SNR of -2 dB for subsequent experiments, as the noise level was still considered comfortable.

For part II, participants had to distinguish a target stimulus made of two 0.2 ms pulses with a variable IPI, from a comparison stimulus made of a single 0.4 ms pulse. This task has been used in vision to measure “two-flash thresholds” (e.g. Lewis, 1967), and is based on the idea that energy should be summed up over time for IPIs shorter than the temporal resolution, a rule known as Bloch’s law (Gorea, 2015). For touch and audition, a similar click-resolution task had also been previously used (Gescheider, 1966). Figure 2, right, displays percent correct on the task as a function of IPI. Performance increased with increasing IPI for Vision. The first IPI that was above chance for Vision was 55 ms ($t(4) = 20.8$, $p < 0.001$). For Touch, all values were close to ceiling and well above chance (e.g. for 55 ms, $t(4) = 51.44$, $p < 0.001$). Even though we did not test Audition, its temporal resolution is known to be an order of magnitude lower than for Vision, as for instance measured with a similar task click-discrimination task (Gescheider, 1966). Based on the pre-test results, a minimum IPI of 60 ms was selected for time patterns used in subsequent experiments.

To summarize, the pre-test served to rule out a potential confound for the Touch condition, by setting the level of an auditory masking noise that prevented the auditory detection of tactile stimuli. This noise was used in the main experiments with the same level for all modalities, even when no masking was required, to equate for its potentially distracting effect. The pre-test also served to select a minimum IPI ensuring that all IPIs of the temporal patterns of the main experiments would be perceptually resolved. We used a two-pulse discrimination thresholds method, which produced results broadly in line with previous studies (Gescheider, 1966; Lewis, 1967). Note that in the main experiments, the value of 60 ms was the enforced minimum IPI, which in effect occurred very rarely in the stochastic patterns. The mean time interval was longer, at 166 ms, which was comfortably above the resolution thresholds estimated here.

3.4.2 Experiment 1

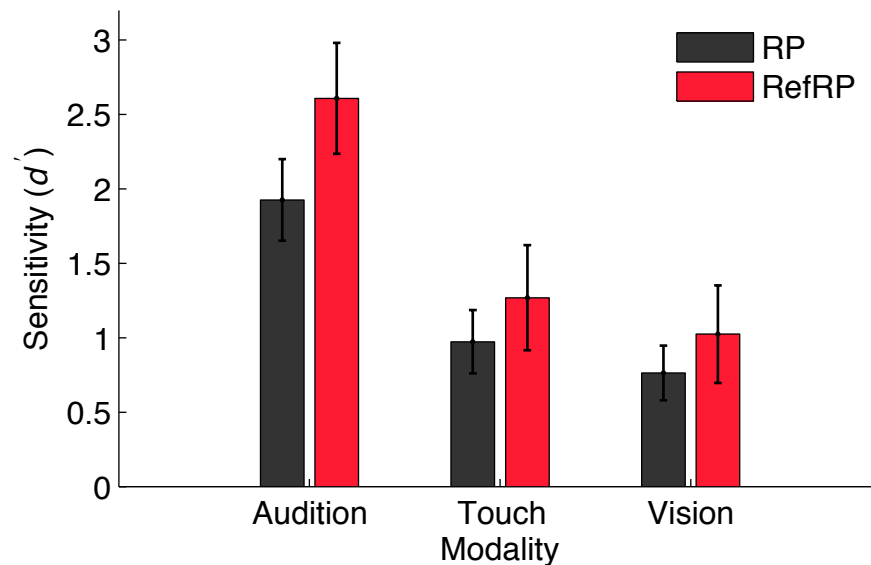


Figure 3-3. Average results for Experiment 1. The discriminability of RP against P and RefRP against P is shown, in d' units, for all three modalities. Results were averaged across participants and error bars represent 95% confidence intervals.

The task required from participants was a within-trial repetition-detection task that has been commonly used to study sensory memory in audition (Guttman & Julesz, 1963; Kaernbach, 2004). Here, the task was applied to three sensory modalities: Audition, Touch, and Vision. In addition, some so-called reference patterns (RefRP stimulus type) re-occurred 20 times throughout an experimental block. Based on results in the auditory modality, we hypothesised that perceptual learning across trials may be observed for such patterns and revealed by an improved performance for RefRP over RP (Agus et al., 2010; Agus & Pressnitzer, 2013; Gold et al., 2014; Kang et al., 2017).

Figure 3-3 displays the average performance, in d' units, for each stimulus type and modality. For Audition, we replicated previous findings, with a relatively high performance on the baseline repetition-detection task for the RP stimulus type. Importantly, performance was better for the RefRP stimulus type, suggestive of perceptual learning. For Touch and Vision, the same effects were observed, with a lower RP baseline performance but always an advantage of RefRP over RP.

A repeated measures ANOVA confirmed that both stimulus type ($F(1,17) = 35.38, p < 0.001, \eta^2_g = 0.11$) and sensory modality ($F(2,34) = 74.28, p < 0.001, \eta^2_g =$

0.52) had a significant effect on performance. An interaction between stimulus type and sensory modality was observed ($F(2,34) = 4.57, p < 0.05, \eta^2_g = 0.03$). Post-hoc tests were conducted to compare RP and RefRP in each modality. Paired t -tests on d' for each modality confirmed the RefRP advantage (Audition: $t(17) = -6.37, p < 0.001$, Vision: $t(17) = -2.15, p < 0.05$, Touch: $t(17) = -2.68, p < 0.05$; all p values uncorrected values but remain significant after Holm-Bonferroni correction). Finally, we tested for the difference between baseline performance on the repetition-detection task across modalities, by performing another repeated-measures ANOVA on d' for the RP stimulus type only. Modality had a significant effect on performance ($F(2,34) = 64.02, p < 0.001, \eta^2_g = 0.57$).

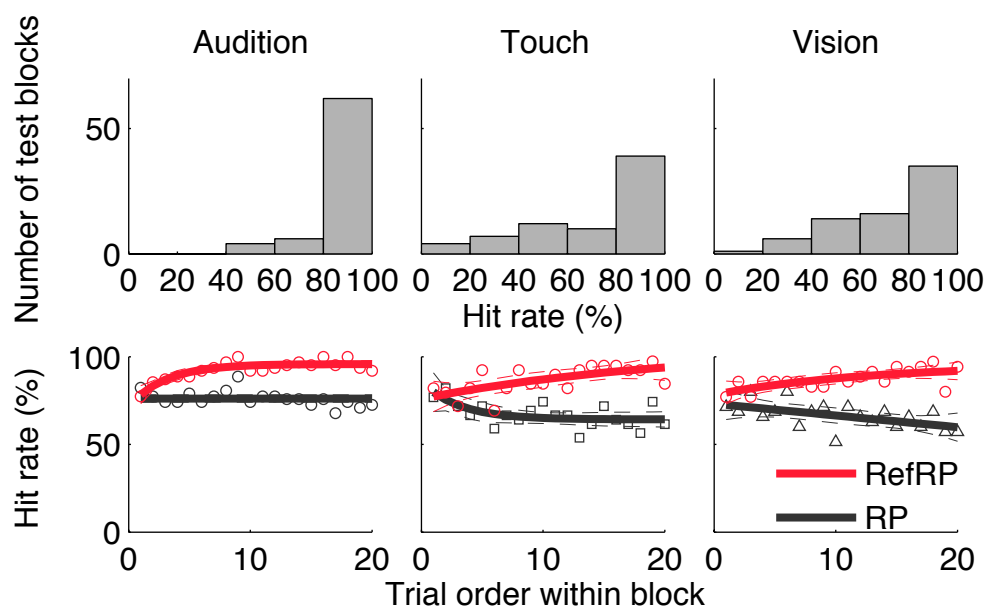


Figure 3-4. The time course of performance for Experiment 1. *Top*: Histograms of the averaged performance over the last 10 presentations of RefRPs, for all experimental blocks and in all three modalities. Good blocks were defined as blocks with 80% or more hit rate over the last 10 presentations. *Bottom*: Hit rate as a function of trial order within the blocks, averaged across participants for good blocks only. The solid lines represent the best fit to the data (see text for details).

Based on previous results in the auditory modality (Agus et al., 2010; Kang et al., 2017), we tested for the presence of “good” and “bad” blocks in the data, defined by the performance on RefRP at the end of a block after learning could have occurred. We computed the average hit rate for the last ten presentations of RefRP, in each test block (Figure 3-4, top). As for previous experiments in the auditory modality, a cluster of

blocks with very high performance (hit rate higher than 80%) was apparent. A one-sample Kolmogorov-Smirnov test confirmed that the distribution of hit-rate over blocks did not match normal distributions (Audition: $D(72) = 0.67$, $p < 0.001$, Touch: $D(72) = 0.54$, $p < 0.001$, Vision: $D(72) = 0.57$, $p < 0.001$). We defined blocks with final hit rates above 80% as good blocks, and the remaining blocks as bad blocks. With such a definition, the bad blocks had an average hit rate at the end of the block of 54% for Audition, 43% for Touch, and 48% for Vision, not different from chance in all cases (all $p > 0.05$ from one-sample t -tests). Good blocks all had a high level of performance, by construction, but they were not evenly distributed across sensory modalities: 86% of all blocks for Audition, 54 % for Touch, and 48% for Vision.

There were two possibilities to explain the high performance in good blocks. It could be that performance gradually improved with exposure to a RefRP pattern, as would be expected if perceptual learning was involved. Or, good blocks could have been the result of a chance selection of temporal patterns producing high performance on the within-trial task even without learning. To distinguish between the two possibilities, we examined performance over time within a block, by averaging all presentations of RP or RefRP in trial bins corresponding to their order of appearance (Figure 3-4, bottom). Two models contrasting either constant performance throughout the block or an exponential change in performance were fit to the data, and their goodness of fit was compared taking into account the number of free parameters (Methods and Agus et al., 2010; Kang et al., 2017). Performance for RefRP increased over time in all modalities (Audition: $F(2, 17) = 28.33$, $p < 0.001$); Touch: $F(2,17) = 6.47$, $p < 0.01$; Vision $F(2,17) = 6.85$, $p < 0.01$). Performance for RP remained constant for Audition ($F(2,17) = 2.77$, $p = 0.09$) but decreased slightly for Touch ($F(2,17) = 4.88$, $p < 0.05$) and Vision ($F(2,17) = 3.59$, $p = 0.05$). Such a decrease in performance over time for RP has been observed in previous auditory experiments and attributed to a criterion effect: as RefRP becomes easier to detect, participants respond less to RP to keep their balance between “yes” and “no” responses (Agus et al., 2010).

MEMORY FOR RANDOM TIME PATTERNS IN AUDITION, TOUCH, AND VISION

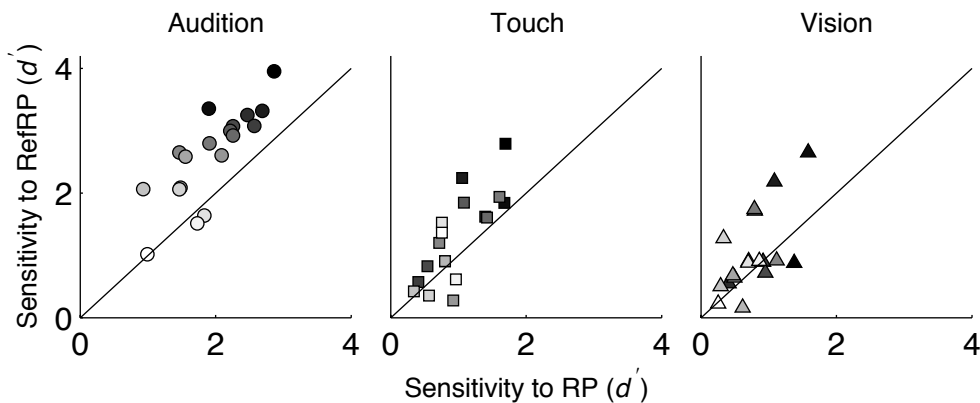


Figure 3-5. Individual data for Experiment 1. Each square represents an individual participant, identified across modalities with the shade of the square. The averaged performance for RefRP is shown as a function of the averaged performance for RP (good and bad blocks averaged together). The diagonal indicates equal performance for both stimulus types. Points above the diagonal are participants with a pattern results consistent with perceptual learning.

Finally, we examined the individual data for each participant, averaged across good and bad blocks. Figure 3-5 displays individual data averaged across blocks, in each modality, with participants identified with marker's shades. For all modalities, there were cases of participants exhibiting a clear advantage of RefRP over RP, with perhaps a higher prevalence for Audition and Touch compared to Vision. This parallels the number of good blocks in each modality.

Table 3-1. Correlation (Pearson correlatin coefficients) for individual d' values across sensory modalities, for stimulus types RefRP, RP, and for the difference in d' between RefRP and RP. Values in bold and italics indicate strong to moderate correlations between a set of modalities (Italics: correlation coefficients (r) above 0.4 with borderline significance $p = 0.06$, bold: $p < 0.05$, p -values all uncorrected).

	Audition - Touch	Audition - Vision	Touch - Vision
RefRP	<i>0.53</i>	<i>0.45</i>	<i>0.45</i>
RP	<i>0.47</i>	<i>0.53</i>	<i>0.50</i>
RefRP - RP	0.10	0.14	-0.04

We further examined whether performance in one modality was predictive of

performance in another modality. Table 3-1 presents the cross-sensory correlations computed across individuals for d' . Correlations across modalities were moderate for RP and RefRP, with uncorrected p -values below or around our significance threshold of $p < 0.05$. We also computed correlations for differences in d' between RefRP and RP, which was one measure of the added benefit of repeated exposure related to perceptual learning. There was no correlation across modalities for the RefRP minus RP d' difference.

3.4.3 Experiment 2

For Experiment 2, the aim was to test whether the learning of a temporal pattern in one sensory modality would transfer to another modality. Because such a test depends on the pattern being learnt in the first place, we chose to present the patterns first in the auditory modality, as this was where most good blocks were observed for Experiment 1. Participants thus started each experimental block by performing the task of the Audition condition of Experiment 1, with P, RP, and RefRP stimulus types and the same number of trials. Then, they switched to the Touch or Vision modality, again with the same task as for Experiment 1. Without informing participants, we contrasted two conditions: a “Transfer” condition where the RefRP in the second sensory modality was the exact same pattern of IPIs as presented in the beginning of the block in the auditory modality; a “Control” condition, for which the RefRPs in the first and second sensory modalities were different patterns of IPIs.

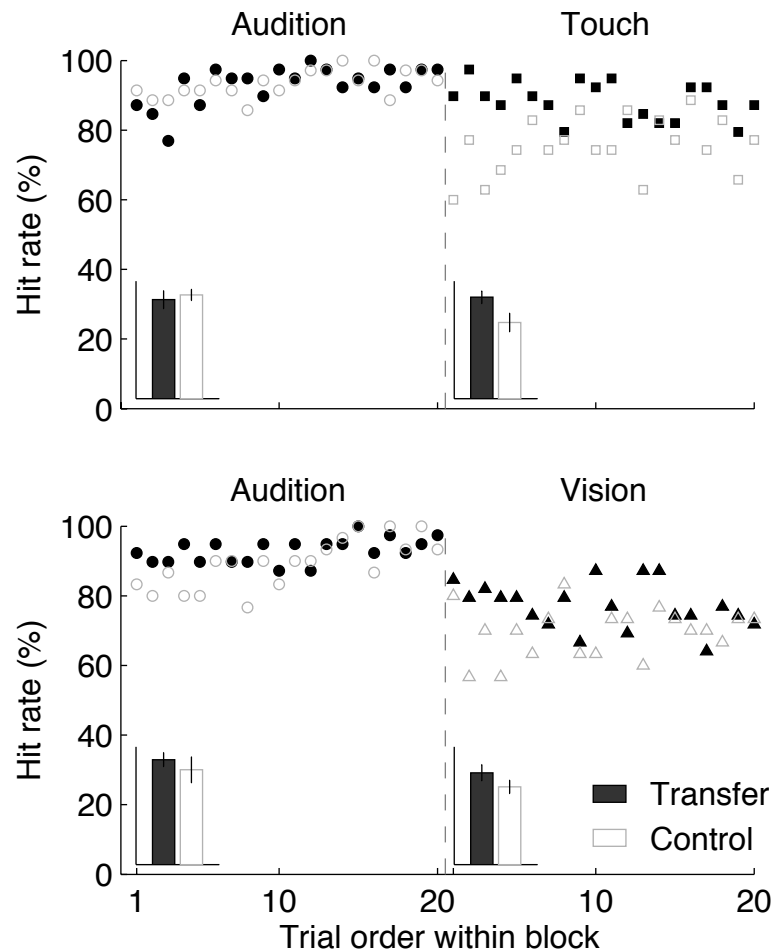


Figure 3-6. Results for Experiment 2. The time course of performance is shown for the Transfer condition (same RefRP patterns before and after the switch) and Control condition (different RefRP patterns before and after the switch). The switch from auditory to either touch or visual presentation is indicated by the vertical dashed line. Only good blocks with evidence of initial learning in the auditory modality were included. The insets show the average performance across participants for the 5 first presentations in each modality. Error bars represent 95% confidence intervals.

Figure 3-6 displays the hit rate for RefRPs in the Transfer and Control conditions according to their trial order in the block (same analysis as Figure 5). For the figure and subsequent analyses, we selected blocks for which there was evidence of learning in the first presentation modality, as else, there would be no point in asking about transfer of learning. This was done by selecting good auditory blocks, defined according to the criterion of Experiment 1. No criterion was applied to the performance in the second modality. In the participant's group that was tested on Audition/Touch, this selected 81% of blocks in the Transfer condition and 73% of blocks in the Control condition. In the group that was tested on Audition/Vision, this selected 81% of blocks in the Transfer

condition and 62% of blocks in the Control condition. One participant did not have any good auditory block in one experimental condition (Transfer Audition/Vision) and was excluded from the analysis.

If there were no transfer of learning across modalities, the performance for RefRPs initially learnt in audition should drop right after the switch to a different sensory modality, to performance levels equivalent to those obtained with novel patterns. Conversely, if there were transfer of learning, performance for RefRPs initially learnt in audition should remain high after the switch, even before any re-learning could happen in the new modality. The pattern of results was consistent with transfer of learning. For Touch, performance right after the switch was essentially the same as performance in the auditory modality after initial learning. For Vision, performance decreased after the switch, but, like for Touch, it remained above what was observed for the novel patterns of the Control condition.

To quantify these observations, we averaged hit rate for the first 5 trials right after the modality switch and performed a mixed-design ANOVA with main factors of condition (Transfer or Control) and modality (Touch or Vision). Condition had a significant effect on performance ($F(1,21) = 37.50, p < 0.001, \eta^2_g = 0.31$). No effect of modality was observed, nor any interaction between condition and modality. Post-hoc t -test were further used to compare performance between Transfer and control, and confirmed an advantage of Transfer for both Touch ($t(11) = -4.89, p < 0.001$) and Vision ($t(10) = -3.78, p < 0.01$).

We finally controlled that the advantage of the Transfer condition after the switch did not result because of a random selection of especially easy RefRP patterns, irrespective of learning (remember that all RefRPs were from good blocks, so with a level of performance matched after possible learning but not before). The same mixed ANOVA was applied to the first 5 trials of auditory presentation in each block. There was no main effect of conditions or modality groups ($p > 0.05$), confirming that all selected RefRPs induced an initially balanced level performance on the repetition-detection task, at least for auditory presentation.

3.5 Discussion

Using a paradigm previously used in audition (Kang et al., 2017), we investigated whether repeated exposure to temporal patterns in audition, touch, or vision would result in similar forms of perceptual learning. We used random sequences of sub-second time intervals, each interval being constrained to be above the temporal resolution limit in all modalities. There were strong similarities in the results in all three modalities. First, the within-trial repetition-detection task that was used, which is a common and powerful tool to probe sensory memory in audition (Guttman & Julesz, 1963; Warren et al., 2001; Kaernbach, 2004), could be performed in touch or vision, as indicated by almost all participants meeting our inclusion criterion during the training period. Second, when some reference patterns re-occurred over several trials in the main experiments, they produced higher performance on the repetition-detection task compared to novel patterns.

This is indicative of rapid perceptual learning, as previously observed in audition (Agus et al., 2010; Kang et al., 2017), but now generalized to three sensory modalities. Importantly, learning was non-supervised. It was also incidental, as participants were not told about the existence of reference patterns, nor did they know that learning some trials could benefit subsequent performance. Moreover, the reference patterns occurred at unpredictable times during the experiment and were presented amongst statistically matched patterns. Finally, whenever learning happened, it was complete in about only ten trials. These are all desirable characteristics for real-life learning of time patterns. There were recent demonstrations of sequence learning for sequences of tactile stimuli consisting of temporal amplitude patterns to the fingertips for human observers or whisker deflections for mice (Bale et al., 2017), also with characteristics similar to auditory incidental memory (Agus et al., 2010). Converging results have been found using the same repetition-detection paradigm for sequences of visual contrast (Gold et al., 2014). Here, we extended these findings with stimuli containing purely temporal information, and the qualitative similarities between modalities became perhaps even more salient.

There were also quantitative differences between modalities, however. In particular, the baseline performance on the within-trial repetition detection task was best in audition, followed by touch and then vision. The amount of learning was also different across modalities, as indicated by a statistical interaction in our analyses. Fewer blocks

led to learning for vision than for audition, which in fact matched the introspective reports of our participants on the difficulty of the task. This suggests an advantage for the learning of time patterns in the auditory modality, followed by touch and then vision. Such a hierarchy would be consistent with a “temporal advantage” for audition, as reported in a broad variety of studies. For instance, audition has an advantage over vision for time interval discrimination (Pasinski et al., 2015) or sequence learning (Nazzaro & Nazzaro, 1970), and touch has an advantage over vision for sequence learning (Manning et al., 1975). Some of these results have been interpreted by positing additional strategies for audition, such as beat-based rhythmic encoding (Povel & Essens, 1985; Patel et al., 2005; Pasinski et al., 2015). However, using moving bars instead of light flashes to delineate time intervals, some evidence has been found for rhythmic processing in vision (Grahn, 2012), without cancelling the overall advantage for audition. Our sequences were not regular, so the present advantage for audition may not be accounted for by rhythmic processing, but rather by more general factors such as different time-constants for sensory memory across modalities (Collier & Logan, 2000).

However, there are also alternative interpretations to consider. Even if our aim was to present purely temporal patterns, there were by necessity choices to be made for the modality-adapted stimuli used to delineate time intervals. These choices were guided by the aim to obtain the most temporally accurate stimulation in all modalities. For instance, a green LED was presented in central vision with moderate background lighting, to preferentially recruit cones and enhance visual temporal acuity (Kalloniatis & Luu, 1995). But it could be that our largely arbitrary choices were not equally effective in all modalities, and that a different outcome could have been obtained with other modality-specific stimuli to delineate time intervals (Grahn, 2012). One possibility to extend the present results would be to forego the constraint of matching sequence statistics across modalities (Marks, 1987), but rather to scale the IPI distributions by *e.g.* discrimination thresholds in each modality (Hansen & Cottrell, 2013). For now, while all the trends in our data would seem highly consistent with the many reports of an auditory advantage for temporal pattern processing, the main result still has to be that incidental learning of complex time patterns was possible and qualitatively similar for audition, touch, and vision.

Such behavioural findings present interesting challenges for current models of neural plasticity. To learn time-localized information, such as quasi-instantaneous frequency profiles in audition, static textures in touch, or still images in vision, there are

many plausible underlying plasticity processes based on the well-established principle of “fire together, wire together” (e.g. Masquelier et al., 2008). In contrast, the time patterns used here contained information that unfolded over more than a second, and they could not be distinguished by summary measures such as average rate. Which other neural strategies could there be to learn such time patterns? Several computational possibilities have been put forward (Hardy & Buonomano, 2016). When neurophysiological evidence is considered, past stimulation has been shown to be reflected in current firing patterns of single neurons in primary auditory cortex (Klompf et al., 2012). At the population level, evidence from the avian brain show a recoding of sequence-like temporal patterns into a rate-place code within secondary auditory cortical areas (Lim et al., 2016). Similar mechanisms for time to place recoding have been proposed for touch (Rossi-Pool et al., 2016) or vision (Nikolić et al., 2009). Importantly, the resulting place code would contain “instantaneous” information and would thus be amenable to well-established neural plasticity processes, applicable to any modality. So, after a modality-specific recoding of the temporal sequences, the similarities across modalities may thus be indicative of common memory mechanisms, consistent with previous suggestions (Pasinski et al., 2015).

A second finding of the present study is the immediate transfer of learning for patterns first presented in the auditory modality to patterns subsequently experienced in the tactile or visual modality. In the data, this was not a subtle effect: the level of performance was maintained almost perfectly when switching from audition to touch. For a switch from audition to vision, performance dropped after the switch, but it still remained above performance for control patterns that had not been learnt in audition. Note that the drop in overall performance may be due to task difficulty associated with the visual presentation, independently of the issue of transfer of learning. Indeed, statistical analyses did not uncover any interaction between modalities and our measure of transfer of learning.

Such a large and symmetrical transfer from audition to either touch or vision has not been reported yet, to the best of our knowledge. Transfer from audition to vision has been observed for time interval discrimination (Bratzke et al., 2012) or rhythm learning (Barakat et al., 2015), although negative evidence also exist (Lapid et al., 2009). There are also examples of transfer from touch to audition (Nagarajan et al., 1998). We will now consider three possible interpretations for the transfer: a recoding of time-interval information into the auditory modality, voluntary or mandatory; the existence of an

amodal representation common to all three modalities; strong and early interactions across modalities.

The first possibility, of a recoding of touch or visual temporal information into an auditory representation, could apply to the transfer data but also to our paradigm in general. Such a recoding could be voluntary and explicit. Imagine that participants knew or guessed that the auditory modality was most appropriate for the repetition-detection task, they may have tried to “auralise” the sequences presented through touch or vision, by imagining a sound for each pulse. The transfer would then not have been across modalities, but across actual and imagined auditory sequences. We considered this possibility when designing the experiments, and the masking noise presented in all modalities was to some extent intended to disrupt an explicit auditory recoding strategy. However, even the presence of noise cannot rule out the imagining of brief sound events. Nevertheless, several elements make an explicit recoding strategy unlikely. First, no participant spontaneously reported such a strategy in post-experiment debriefs. Second, there were differences in overall performance between touch and vision, even though both modalities should present an equal opportunity for an explicit recoding strategy. Third, sequences were fast and unpredictable, making intentional recoding more difficult. Finally, for Experiment 1, we performed an additional analysis comparing the performance of participants who started with auditory presentation, to the performance of those who started with either touch or vision, as starting with the easier auditory modality may have favored an explicit recoding strategy in later blocks. To do so, we added a factor of “audition first (yes/no)” to the ANOVA used to analyze the results of Experiment 1. The main effect for this factor was not significant ($F(1,16) = 0.32, p > 0.05$). So, even if this cannot be formally ruled out, we would argue that an explicit recoding strategy was unlikely to explain the transfer results and general similarity of results across modalities.

A more interesting possibility is that recoding may have occurred involuntarily, perhaps as the default mode for processing timing sequences irrespective of the modality of entry. There is behavioral support for such an idea (Guttman et al., 2005). In this study, a temporal sequence discrimination task was performed in the visual modality, in the presence of task-irrelevant information in the auditory or visual modality. Only distractors in the auditory modality had a negative impact on visual performance, which was interpreted as a mandatory recoding of visual temporal patterns into an auditory code to perform the sequence discrimination task (see however Grahn, 2012).

Neurophysiological evidence along the same lines showed a decrease in performance for either auditory or visual time interval discrimination after TMS inactivation of auditory regions (Kanai et al., 2011). The transfer and similarities observed across modalities in our behavioral tasks could thus be consistent with a mandatory recoding of visual or tactile information into an auditory-like representation.

A second possible kind of interpretation would posit a common representation for time sequences as well, but an amodal rather than auditory representation. Such a representation could be closely related to motor networks. Plastic changes associated with perceptual learning of visual time intervals have been observed in motor parts of the cortex (Buetti et al., 2012). Priming effects between auditory and visual rhythms may also be mediated by motor networks (Grahn et al., 2011). While it is beyond the scope of present paper to review the plausibility of an amodal representation of time (Ivry & Schlerf, 2008; van Wassenhove, 2009; Grondin, 2014), we note that such a representation would certainly be sufficient to explain the transfer effects we observed. However, it may not be easily able to explain the differences in absolute level of performance between the touch and visual modalities observed in the transfer experiment, nor the quantitative advantage of audition in the first experiment.

Finally, a third possible interpretation is that the encoding of time sequences was modality-specific, but that there were strong interactions between modalities before or at the level where the neural learning of time sequences occurred. Based on the evidence reviewed in the Introduction, timing information may be recoded into a place code in secondary sensory areas (Lim et al., 2016; Rossi-Pool et al., 2016). Anatomical evidence for earlier cross-modal connections exist for audition, touch, and vision (Cappe & Barone, 2005), with perhaps especially strong links between audition and touch (Occelli et al., 2011). About this later point, it has been shown behaviorally that tactile stimulation could be masked by auditory stimulation in a frequency-specific way (Yau et al., 2009), suggesting interactions at levels where frequency channels exist, such as sensory cortex. An especially strong connection between audition and touch (Occelli et al., 2011; Saal et al., 2016) would be consistent with the pattern of results we observed. Speculatively, the temporal patterning of multisensory events is often correlated across modalities, so it may indeed make functional sense to have cross-talk between modalities at several stages of processing to facilitate the learning of time sequences.

Conclusion and Perspectives

In a series of behavioral experiments, we have shown rapid perceptual learning for time patterns in audition, touch, or vision. This extends previous results in the auditory modality (Kang et al., 2017) and complement recent evidence highlighting the importance of sequence processing in all three modalities (Gold et al., 2014; Bale et al., 2017). We have also shown that learning a time pattern in audition transferred immediately to touch or vision. This was interpreted by either a mandatory recoding of temporal information into an auditory representation (Guttman et al., 2005), by an amodal representation for temporal sequence (Ivry & Schlerf, 2008), or by a distributed encoding of time combined with early interactions between modalities and common memory mechanisms (Cappe & Barone, 2005; Pasinski et al., 2015).

Learning was described as incidental, because it was unsupervised and occurred during the performance of an ancillary task. We did not test whether participants became aware of their learning, but previous studies in audition showed that the present paradigm led to EEG correlates of memory traces even when attention was diverted from the sounds to be learnt (Andrillon et al., 2015), and to behavioral and EEG correlates of learning for sounds that were heard during light sleep (Andrillon et al., 2017). Extrapolating from these findings, it is thus possible that re-occurring time patterns always leave a sensory memory trace, whatever the modality. What would be the functional benefit of such traces? Their use to encode and recognize temporal patterns obvious, but they could also be involved in more general sensory prediction processes, as suggested by neural evidence in visual cortex (Gavornik & Bear, 2014). In any case, the present behavioral data raises the intriguing possibility that a rapid incidental learning of time patterns is a general feature of sensory systems.

Acknowledgments

We would like to thank Morina Bringezu for running pilot versions of the auditory and visual experiments. We also thank Vincent Hayward for his guidance on the design of the tactile stimulation box. This work was funded by the ERC Grant ‘ADAM’ No. 295603, the ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL*.

Chapter 4

Memory for random time patterns of electrical pulses presented through cochlear implant

Parts of the present work have been presented as:

Kang, H., Macherey, O., Roman, S., & Pressnitzer, D. (2016). Auditory Perceptual Learning in the Time Domain for Cochlear Implant Users. Congrès Francais d'Acoustiques, Le Mans, France (*Talk*).

Kang, H., Macherey, O., & Pressnitzer, D. (2016). Auditory Learning of Temporal Information for Normal-Hearing Listeners and Cochlear Implant Users. 39th MidWinter Meeting of the Association for Research in Otolaryngology, San Diego, USA (*Poster*).

4.1 Abstract

Cochlear implants (CIs) are neuroprosthetic devices used to electrically stimulate the cochlea of patients with severe hearing loss. In a current typical device, a limited number of electrodes stimulate different parts of the cochlea corresponding to different frequency regions, with an electrical signal shaped to convey the temporal amplitude envelope of the external signal in the corresponding frequency band. Because of the limited number of electrodes and even more limited number of independent frequency channels, CI users have to rely heavily on temporal information to perform everyday auditory tasks such as speech intelligibility. Accordingly, several studies have shown that temporal processing as measured by discrimination tasks is broadly preserved in CI users, at least compared to the sharp degradation observed in spectral discrimination tasks (Moore & Carlyon, 2005). In the present chapter, we investigated another aspect of temporal processing beyond immediate discrimination: the ability to learn complex temporal patterns. We adapted the

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

paradigm of Kang et al. (2017), presented in Chapter 2. Based on our previous results, we first chose 2 representative average pulse rates where learning was observed for normal hearing (NH) listeners, 10 Hz and 40 Hz. We compared performance for a NH control group, stimulated acoustically, and a CI group, using direct electrical stimulation of the cochlea through the research interface of the implant to convey the complex time patterns. Furthermore, for CI listeners we added a 300 Hz pulse train rate that would not easily be restricted to purely temporal cues across the whole cochlea for NH listeners. We observed qualitatively similar results for the CI and NH groups on rates that were presented in both groups, with rapid learning of electrical time patterns in CI listeners. Evidence of learning was also found in the CI group for the fastest rate. These results have clinical relevance, as they show that the plasticity processes involved in rapid perceptual learning may be largely preserved in CI listeners. Moreover, on a fundamental level, the results confirm that pure timing cues are amenable to perceptual learning, even in a timing range that is not normally encountered by the healthy auditory system.

4.2 Introduction

The ability to understand complex acoustic scenes depends in a large part on memory processes, which enable listeners to recognise the sound sources in their environment thanks to past experience. For normal-hearing listeners, these processes are likely to occur throughout development as well as in adulthood, such as when for instance learning to recognize the voice of a new acquaintance. As we have repeatedly mentioned during this thesis, there is now evidence for the availability of remarkably fast and efficient learning processes that could subservise this ability even in adulthood (Agus et al., 2010; Kang, Agus, & Pressnitzer, 2017).

Learning novel sounds is perhaps an even more acute problem for CI users. Indeed, for post-lingually implanted adults, the switch from acoustic hearing to electric hearing involves a complete re-learning of the meaning of all sounds in their environment, because the cues available with a cochlear implants are drastically different from those available acoustically. Moreover, the learning of temporal information is also crucial for CI listeners. Because of limits in the current CI devices, both in the number of physical electrodes (up to 22 in modern implants) and even more acutely by the difficulty to achieve focal electrical stimulation in the conductive cochlear fluids, spectral resolution of CIs is limited (Macherey & Carlyon, 2014). Instead, CI users have to rely on temporal cues to achieve good performance on tasks such as speech intelligibility (Shannon et al., 1995). Here, we tested the characteristics of perceptual learning of temporal patterns for CI listeners, as this would be a process useful to them in the re-learning of their acoustic environment through temporal cues.

Previous studies have shown that temporal processing is generally preserved for CI listeners, at least when measured with discrimination tasks. Experiments measuring detection thresholds at different temporal rates (Shannon, 1985), pitch tasks (Zeng, 2002; see Moore & Carlyon, 2005 for a review), rate discrimination thresholds (Gaudrain, Deeks, & Carlyon, 2017) showed that pulse rate perception was preserved up to about 300 Hz. Only few studies studied temporal perception in CI implants with irregular time patterns, such as the ones we used in the present work. Kong et al. (2004) measured the musical abilities of CI users evaluated by pitch and rhythmic tasks. While CI users showed large deficits on the pitch tasks, their performance was comparable to NH listeners on the rhythmic tasks. Gaudrain, Deeks, and Carylton (2017) studied fully

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

random sequences, even within a rhythmic structure, in a jitter-detection task in addition to the more classic rate discrimination task. They found similar performance between CI and NH listeners on the jitter detection task for average pulse rates of 100 Hz, and slightly degraded performance for the CI users for average pulse rates of 300 pulse-per-second (pps). Interestingly, no correlation was observed between performance on rate discrimination and jitter detection, implying that the classic measures of rate discrimination may only give a partial view of all of the processes involved in temporal perception with a CI.

Given that CI users may have a perceptual representation of random temporal patterns sufficient to support immediate discrimination, here we asked whether the neural processes required for learning such patterns was preserved. From previous studies (Andrillon et al., 2015; Luo et al., 2013; Kumar et al., 2014), it has been hypothesised that the rapid learning of complex sounds recruited at least secondary auditory areas. These areas are known to undergo plastic processes during the sound-deprivation period that accompanies deafness (Ponton, Moore, & Eggermont, 1999). It is therefore unknown whether CI users will have retained the neural bases required for rapid perceptual learning. Answering such a question has obvious clinical relevance.

Furthermore, CI users present a unique opportunity to study temporal processing in the auditory system. By using the research interface of CIs, it is possible to bypass the standard sound processing algorithms designed to mimic typical auditory function. Instead, it is possible to deliver electric pulses at precise time in the whole electrode array, limiting to purely temporal cues the information available to further stages of auditory processing. And, because the timing information sent to electrodes is not limited by acoustic constraints or biophysical processes in the cochlea, it is also possible to study temporal perception for patterns that would not be normally available in NH listeners. For instance, here we decided in some conditions to stimulate the electrode array with fast temporal information, at an average rate of 300 pps. Such fast rates, if presented acoustically, would be converted to spectral information in apical regions the healthy cochlea (Shackleton & Carlyon, 1994). If those temporal patterns could still be learnt, this would suggest that the rapid learning of temporal patterns recruits fundamental and ubiquitous principles of neural plasticity.

To measure the learning of electrical time patterns, we adapted the repetition detection paradigm used in previous chapters (Agus et al., 2010; Kang et al., 2017).

Instead of using audio clicks to delineate time intervals, we used electrical pulses transmitted quasi-simultaneously to 4 electrodes of the CI, covering all of the useful frequency range of the device. A control group using NH listeners was also tested with the usual acoustic stimulation, but using the temporal patterns parameters used in the CI group. Based on our previous results, we chose 2 representative average pulse rates, 10 Hz and 40 Hz, to test both CI and NH listeners. For the CI listeners only, we added a 300 Hz pulse train rate that would not easily be restricted to purely temporal cues as for NH listeners.

4.3 Experiment

Methods

Participants

Ten CI listeners (age range: 37 – 77, $M = 55$, $SD = 15.6$, 4 male) took part in the experiment. CI listeners were recruited in the Marseille area from a database of CI patients regularly tested in the laboratory of O. Macherey (Laboratoire de Mécanique et d'Acoustique, UPR 7051). All CI listeners were recipients of a Med-El CI. Seven CI listeners participated in two test sessions, using different pulse rates (both slow and fast), and three CI listeners participated in only a single test session, two of them for slow-rate pulses and one of them for fast-rate pulses. In total, nine CI listeners participated to the test with slow-rate pulse trains and eight CI listeners participated to the test with fast-rate pulse trains.

Nine NH listeners were recruited in Parisian area as in our previous experiments. There was no attempt to match the CI and NH group in age. The age for the NH group was between 18 – 35 ($M = 22.5$, $SD = 0.7$, 2 male). All NH listeners passed an audiometry test measuring thresholds at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. All participants had threshold smaller or equal to 20 dB HL. All participants provided written consent and were remunerated for their participation (NH listeners) and travel (CI listeners). The experiment with CI listeners was ethically approved by the local research ethics committee (Eudract 2012-A00438-35) and the experiment with NH listeners was ethically approved by the ethical committee of U. Paris Descartes, France (CERES, IRB: 20154000001072).

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

Stimuli and Task

Stimuli consisted of sequences of pulses with random inter-pulse-intervals (IPIs) drawn from a Poisson distribution with a refractory period of 10 ms for 10 Hz and 40 Hz, and 0.3 ms for 300 Hz. Three conditions were tested. Rates of 10 Hz and 40 Hz were used for CI and NH listeners, and an additional rate at 300 Hz for the CI group.

For CI listeners, these pulses were conveyed as electrical pulses through the research interface of the implant. Biphasic pulses were used with phase duration of 45 μ s in monopolar mode of stimulation (MPI -2 negative, amplitude range: 2, Imp type: Pulsar, pulse type: B). Each pulse sequence lasted for 2 s. For the 10 Hz rate, a pulse train with IPIs drawn from a Poisson distribution with refractory period was first computed. Then, the same pulse train was transmitted to 4 different electrodes covering the whole frequency range, with approximately regular spacing between electrodes. Electrode 1, 4, 7, and 10 were chosen for all participants except three, for whom some of these electrodes were switched off. For these participants, electrodes 2, 4, 7, and 10 or 1, 4, 6, and 8 or 1, 4, 7, and 9 were chosen. The sequences were played quasi-simultaneously on all electrodes, with only a 100- μ s delay between electrodes to avoid direct electrical field summation. For the second condition, with an overall average rate of 40 Hz, four different pulse trains with an average rate of 10 Hz were created and then aggregated into one single pulse train. This pulse train was then transmitted to the same 4 electrodes quasi-simultaneously with a 100- μ s delay across electrodes. This generation method was motivated by another set of conditions where we interleaved different 10 Hz sequences on the 4 electrodes. However, the data analysis being incomplete on those conditions, they will not be reported in the present document. For the 300 Hz rate, the same generation procedure as in the 10 Hz rate condition was adopted, with a different rate parameter for the Poisson distribution.

For the NH group, only 10 Hz and 40 Hz were used. Sequences with IPIs corresponding to pulse trains of 10 Hz and 40 Hz were created in the same way as for the CI group. Pulses were presented acoustically. Each pulse duration was 50 μ s. Pulse trains were high-pass filtered using an 8th Butterworth filter with a 1 kHz filter cut-off. No masking noise was used in this experiment. The sound level was kept at overall average 65 dB SPL A-weighted, for both rates.

As in all previous experiments, there were 3 stimulus types. For the first stimulus type, the pulse trains consisted of fully random IPIs for a total duration of 2 s (random

pulse trains, P). In the second stimulus type, a random sequence of duration 1 s was immediately repeated identically, to create a within-sequence repetition (repeated pulse trains, RP). In the third stimulus type, we kept the same RP sequence of random IPIs for several trials within an experimental block (reference repeated pulse trains, RefRP). Different types of stimuli were presented randomly during an experimental block, with the only constraint that no two successive trials should contain RefRPs. The task as explained to listeners was to report within-trial repetitions.

Procedure and Apparatus

CI listeners were tested in a quiet room. They received direct stimulation through the Med-El device and the RIB2 interface. The experiment was programmed in Matlab. Slow-rate pulse trains (10 Hz and 40 Hz) and fast-rate pulse trains (300 Hz) were tested in different sessions. First, for the test with the slow-rate pulse trains, the presentation level of the sequences was determined by measuring the most comfortable sound level (MCL) of pulse trains at the rate of 40 Hz. MCLs were determined by asking participants to indicate the sound level of stimuli with varying electrical amplitude on an 11-point clinical loudness scale with 7 as the MCL. Once the MCL for individual electrodes were collected, stimuli using quasi-simultaneously presentation on the four selected electrodes were presented to find the MCL. We finally checked that pulse trains at the rate of 10 Hz were well audible and the same levels were used in all conditions.

After the MCL adjustment, the experiment proceeded with a training session aiming at providing listeners with examples of the stimuli and task. The training session contained 4 blocks, starting with easy conditions with 10 within-trials repetitions down to 4, 3, and 2 repetitions of 1 s pulse train sequences (Kang et al., 2017). Note that no RefRP was included in the training session. Listeners received visual feedback during the training session. Then, the main experiment was run. It included 4 blocks for each rate condition, without feedback. The order of conditions was randomised across listeners, and 2 test blocks were presented after the full training session for each condition. Another 2 blocks for each condition were run after one additional 2-repetitions training block, in the reverse order for each listener to minimise ordering effects.

For NH listeners, the procedure was the same except that this group did not require the MCL adjustment. Rather, they received the stimuli at a fixed level for all participants. NH listeners were tested in a sound-shielded booth (Industrial Acoustics). The experiment for CI listeners lasted about 2.5 – 3 hrs and for NH listeners the experiment

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

lasted about 2 hrs.

A separate experimental session with 300 Hz pulse trains was run for CI listeners. Invited participants went through the same procedure of adjusting the sound level with 300 Hz pulse trains. They then received a training session with 10, 4, and 3 repetitions, and five blocks of test with 3 repetitions followed. No test with 2 repetitions was done for 300 Hz pulse trains due to the level of task difficulty with the high-rate pulse trains.

Statistical analyses

Responses “yes” in the repetition-detection task for RP and RefRP were considered as hits responses while “yes” for P were considered false alarms. Hits and false alarms were then used to compute the sensitivity index d' from signal detection theory. A repeated-measures analysis of variance (ANOVA) was applied to the d' values, with stimulus type (RP and RefRP) and rate condition (10 Hz and 40 Hz) as within subject factors and participants as random factors, for each group (CI and NH) separately. Mauchly’s test for sphericity confirmed that the data meet the assumptions required for ANOVA. F values and p -values meeting the level of $p < 0.05$ were considered as significant differences. For the repeated measure ANOVA, generalised eta squared (η^2_g) was also reported to indicate effect sizes. Post-hoc paired t -tests were conducted with multiple comparison corrections using the Holm-Bonferroni method. The Pearson correlation coefficient between age and performance was also computed to test whether the group differences may be explained by age. Lastly, a paired t -test was applied to d' values for the 300 Hz test session performed by CI listeners only.

4.4 Results

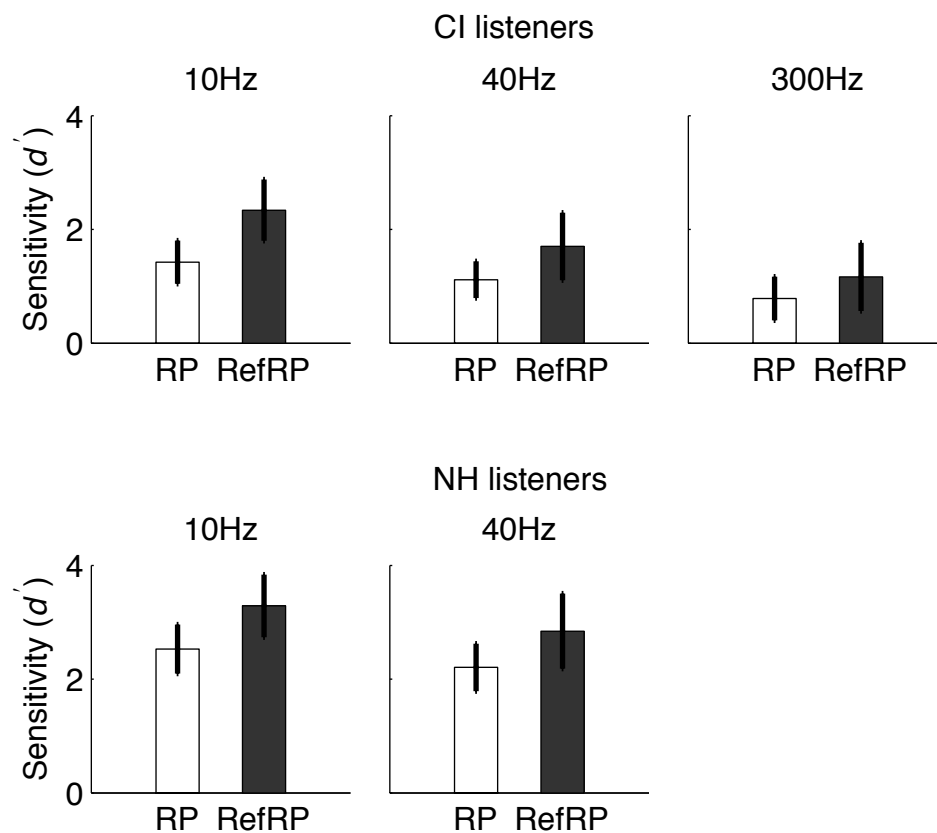


Figure 4-1. Performance on the within-trial repetition detection task measured with the sensitivity index d' . Performance is shown for the different stimulus types (RP, white, and RefRP, black) and displayed separately for CI listeners (top) and NH listeners (bottom). Error bars represent 95% confidence intervals.

The results for both stimulus types and participants group are shown in Figure 4-1. As can be seen from the overall trend, qualitatively similar results were observed in all cases: RefRP produced better performance than RP, suggestive of perceptual learning (Kang et al., 2017). The statistical analyses confirmed this observation. For CI listeners in the slow-rate session (10 Hz and 40 Hz), the repeated measures ANOVA confirmed that listeners showed better repetition detection for RefRP compared to RP ($F(1,8) = 44.76$, $p < 0.001$, $\eta^2_g = 0.37$). Performance also differed across rate conditions ($F(2,16) = 4.45$, $p < 0.05$, $\eta^2_g = 0.12$), with poorer performance for the 40 Hz condition. No interaction between stimulus type and condition was found. Further post-hoc t -tests showed significant differences between RP and RefRP stimulus types for each rate condition (10

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

Hz: $t(8) = -4.59, p < 0.01$, 40 Hz: $t(8) = -3.80, p < 0.01$).

The results of NH listeners were similar. The ANOVA showed an effect of stimulus type ($F(1,8) = 54.08, p < 0.001, \eta^2_g = 0.18$) and an effect of rate condition ($F(2,16) = 6.23, p < 0.01, \eta^2_g = 0.06$) with no interaction between stimulus type and conditions. Post-hoc paired t -tests further confirmed the effect of stimulus type for each rate condition (10 Hz: $t(8) = -6.16, p < 0.001$, 40 Hz: $t(8) = -5.08, p < 0.001$).

Another set of a mixed-design ANOVA combining both groups of listeners was performed. In this analysis, a significant effect of group on performance was observed ($F(1,16) = 23.97, p < 0.001$). However, there was no interaction between group with either stimulus type or rate condition. This indicates that, even if the baseline performance was better in NH listeners compared to CI listeners, there were similar effects of perceptual learning as titrated by the difference between RefRP and RP.

Finally, because our groups were not matched in age, we computed the correlation coefficient between age and d' for each condition. It was -0.43 and -0.47 respectively for rates of 10 Hz and 40 Hz.

Lastly, for the CI listeners who performed the task in the fast-rate condition we used a simple paired t -test to compare performance on RP and RefRP. This confirmed that RefRP produced high performance than RP (300 Hz: $t(7) = -2.9, p < 0.05$).

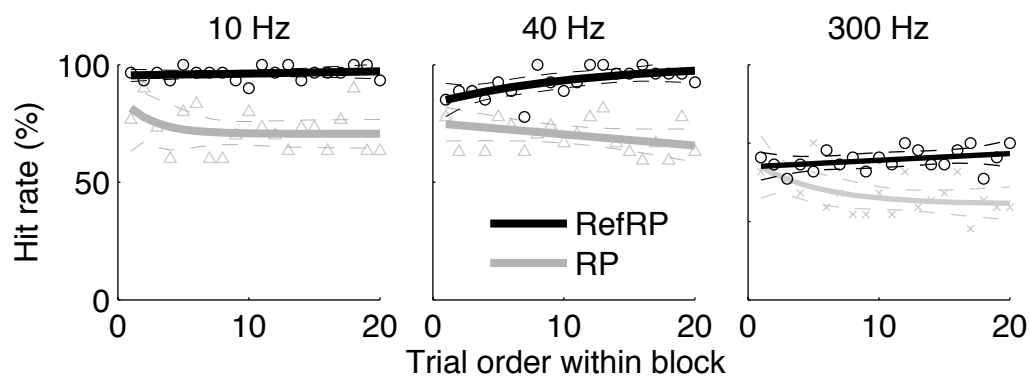


Figure 4-2. Average hit rate of "good blocks" based on each trial order for RP (grey triangles) and RefRP (black circles) in each condition for CI listeners.

As was done in Chapters 2 and 3, we further examined whether the superior performance for RefRP was due to a performance change during the block, indicative of learning, or rather to a chance selection of particularly easy RefRPs. This was done by

computing the hit rates over trials in each condition (see Chapter 2, section 2.3, 4. *Data analysis*). We first computed the average hit rates in the RefRP stimulus type condition over the last 10 trials of a block. For CI listeners, the average hit rates during the last 10 trials for RefRPs were not normally distributed, as shown by a one-sample Kolmogorov-Smirnov test (10 Hz: $D(36) = 0.62$, $p < 0.001$, 40 Hz: $D(36) = 0.60$, $p < 0.001$, 300 Hz: $D(40) = 0.57$, $p < 0.001$). As before, we set a criterion of hit rate at or above 80% to identify "good blocks" where learning may have occurred (Kang et al., 2017). This procedure led to 83, 75, and 55% good blocks in the 10 Hz, 40 Hz, and 300 Hz conditions, respectively. Figure 4-2 shows the time-course of hit rate for RP and RefRP stimulus types over trial order. Again as before, we compared fitted models with either flat performance or exponential change in performance over trials (Kang et al., 2017). This showed that only the RefRP hit rates for 40 Hz gradually increased (40 Hz: $F(2,17) = 5.95$, $p < 0.05$) whereas the hit rates of RP stayed relatively constant (all $p > 0.05$). We did not observe an increase in hit-rate increase for RefRP in the 10Hz and 300 Hz conditions, possibly due to a ceiling effect. For 300 Hz, the hit rate showed a trend of for an increase although this did not meet our significance level ($F(2,17) = 3.11$, $p = 0.071$).

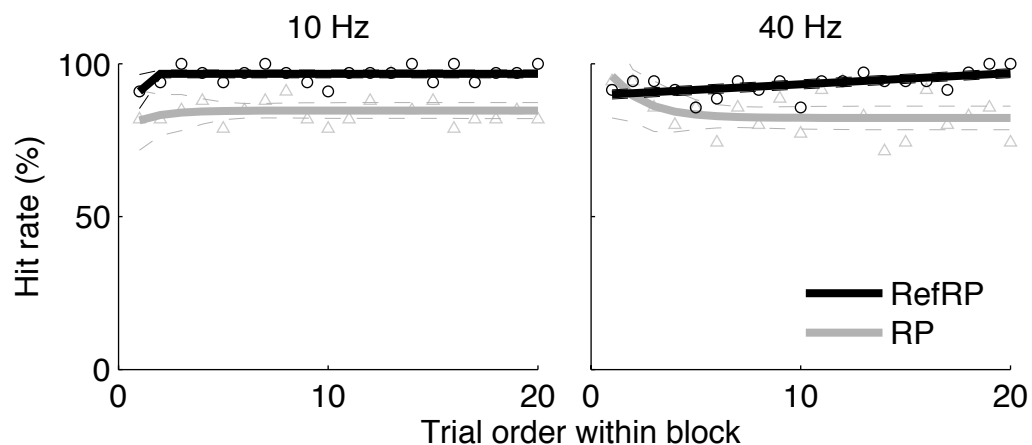


Figure 4-3. Average hit rate of "good blocks" based on each trial order for RP (grey triangles) and RefRP (black circles) in each condition for NH listeners.

The same analysis of performance over time in the NH group is shown in Figure 4-3, with a similar pattern. A one-sample Kolmogorov-Smirnov test showed that average hit rate for RefRPs over the last 10 trials did not conform to a normal distribution (10 Hz:

MEMORY FOR RANDOM TIME PATTERNS OF ELECTRICAL PULSES PRESENTED THROUGH COCHLEAR IMPLANT

$D(36) = 0.76, p < 0.001$, 40 Hz: $D(36) = 0.76, p < 0.001$). Almost all blocks, regardless of conditions, showed high hit-rates. The criterion for good blocks was maintained at hit-rate at or above 80% over the last 10 trials. This produced 92% and 97% good blocks in the 10 Hz and 40 Hz conditions, respectively. With the model fit, we only found a significant improvement over trials for RefRP in the 40Hz condition ($F(2,17) = 3.96, p < 0.05$). No such improvement was seen for the 10Hz condition, probably again due to a ceiling effect. Performance for RP did not show evidence for a gradual change either (all $p > 0.05$).

4.5 Discussion

The present study explored the possibility of rapid perceptual learning of time patterns in CI listeners, and compared their performance to a NH group with matched stimulus characteristics, but different average age (the NH group was younger on average). We found evidence for rapid perceptual learning in CI listeners: the performance for the RefRP condition was always above that of RP, for all rate conditions including the fast one, which would include cues unlikely to occur with acoustically-presented sounds and normal hearing. Remarkably, the benefit of re-occurring RefRPs over novel RPs was similar for the CI and the NH group, even if baseline performance on the RP stimulus type varied. This would suggest largely preserved abilities to learn temporal information in CI listeners, and even an ability to learn temporal patterns with cues specific to electric hearing. These findings would thus extend previous studies demonstrating accurate perception of time-related information for CI listeners (Gfeller & Lansing, 1991; Gfeller, Woodworth, Robin, Witt, & Knutson, 1997; Kong, Cruz, Jones, & Zeng, 2004; Shannon, 1985).

There are some caveats to this interpretation. For CI listeners, 2 rate conditions out of 3 failed to show a significant increase in RefRP performance over time, which is another one of our criterion to ascertain rapid perceptual learning (Agus et al., 2010; Kang et al., 2017). So, in a conservative perspective, we have only fully demonstrated that perceptual learning occurs with electrical pulse rates at a 40 Hz average rate. However, several aspects of the data suggest that our failure to demonstrate changes in performance over time was due to methodological issues and not alternate interpretations.

First, and much to our surprise, performance on the task was quite high even for the CI group. It is noticeable that we failed to find evidence for changes in performance also in the NH group in this condition, whereas previous experiments showed such an increase for similar stimuli (Kang et al., 2017). It would thus be useful to increase the number of participants in future extensions of the study to get a better estimate of the evolution of performance over time. Furthermore, a follow-up study with only an apical stimulation with high-rate pulse trains, the range where no equivalent acoustic stimulation in NH listeners is available, would also be useful to explore the effect that is specific to electric hearing.

Chapter 5

Pupil dynamics as an indicator for auditory memory

Parts of the present work will be presented as:

Kang, H., & Pressnitzer, D. (2018). Pupil Dynamics during Auditory Incidental Learning. 41th MidWinter Meeting of the Association for research in Otolaryngology (ARO), San Diego, USA (*Poster*).

5.1 Abstract

Throughout this thesis, we have argued that auditory memory may be a general by-product of auditory processing, as suggested by a series of studies showing rapid incidental learning for tokens of white noise after a few repeated presentations (Agus et al., 2010) or by rapid incidental learning of time patterns in the same experimental conditions (Kang et al., 2017). Correlates of the phenomenon have been found using standard brain imaging techniques such as EEG (Andrillon et al., 2015), MEG (Luo et al., 2013), or fMRI (Kumar et al., 2014). However, the behavioural studies require time-consuming lab experiments, and the imaging studies additionally involve costly equipment requiring further experimental efforts to set-up. In this chapter, we explore a new tool that may prove a useful complement to existing techniques for future investigations of the behavioural characteristics and neural bases of auditory learning. We tracked pupil size during the incidental learning of noise tokens, to test for another physiological index that had been previously shown to be modulated during auditory learning (Goldinger & Papesh, 2012).

Unlike in previous chapters, here we used a standard memory for noise paradigm, with Gaussian noise stimuli. Noise sequences lasting 1.5 s were presented to adult participants. Each sequence was either 1.5 s of purely random noise (Noise, N), or

consisted of 3 seamless repeats of the same, 0.5 s random noise token, generated afresh for each trial (Repeated Noise, RN). A last condition was like RN, but with the same noise token used in all trials of an experimental block (Reference RN, RefRN). Participants were instructed to report within-trials repeats. While they performed the task, we used a low-cost eye tracker (Eyetribe, acquisition at 60 Hz sampling rate) to record pupil size. Pupil diameter was recorded and baseline-corrected for each trial, with 5s of inter-trial silence to allow pupil size to return to baseline.

Behavioural results replicated previous findings. A large advantage for RefRNs compared to RN was found at the end of each block, indicating perceptual learning of RefRN tokens that had been heard multiple times during a block. The novel result was that pupil dilation was strongly modulated by stimulus condition. Pupil dilation was equivalent for all conditions at the beginning of a block, but was markedly reduced at the end of a block for RefRN only. Thus, sounds that had been learnt induced less pupil dilation than acoustically similar but novel sounds.

The effect of auditory learning on pupil dilation was opposite to what was previously observed during intentional learning (Papesh, Goldinger, & Hout, 2012). A possible interpretation is that pupil dilation here reflected listening effort (e.g. Koelewijn et al., 2015), as RefRNs were associated to near-perfect task performance after incidental learning. Future experiments could disentangle memory effects from listening effort using a similar paradigm, but without a task (e.g. Andriillon et al., 2015).

5.2 Introduction

There are many further experimental manipulations that could be tested to investigate the detailed behavioural characteristics underlying the rapid and incidental learning of complex sounds observed in previous chapters, for instance to map the decay characteristics of the memory traces in addition to the time course of their emergence. As we reviewed in the general Introduction to the thesis, there are also studies that reported correlates of such auditory memories, using EEG, MEG, and fMRI (Andrillon et al., 2015; Kumar et al., 2014; Luo et al., 2013), but clearly further correlates to specify the neural mechanisms involved would be useful. However, both the behavioural and imaging techniques are costly and require time-consuming experiments. For future studies on incidental auditory memory, it may be useful to have a correlate of learning that is easily collected and that may appear even without asking an explicit response from the participant, to speed up data collection. To explore such a possible correlate, we used a simple and cheap (100 USD) eye tracker to measure pupil size during rapid incidental learning of complex sounds.

The idea stems from previous reports relating pupil dilation and memory, both in the auditory and in the visual modality. In the visual modality, Kafkas and Montaldi (2011) used a "remember/know memory paradigm" to measure the effect on pupil size changes of different levels of memory encoding. Participants were first asked to simply look at a set of man-made or natural visual items without knowing the following task. After a break, another set of visual items was presented, which included some of the previously-seen items. Participants were asked to respond about the familiarity of the item: did they recognise it from the first part of the experiment, or were they familiar with the item from past experience. The range of response choices included three different familiarity levels (*weakly familiar* to *very familiar*), as well as *new* and *recollected*. The aim was to disentangle the effect of memory and familiarity on pupil size. The authors observed a decrease of the peak of pupil size with stronger familiarity and better recollection of the visual stimuli. Furthermore, recollected stimuli produced smaller pupil dilation than familiar stimuli.

Again in the visual modality, Naber, Frassle, Rutishauer, and Einhauser (2013) studied pupil size as an index for memory retrieval as well as memory formation for natural images. The encoding phase of their experiment was either explicit, by instructing

participants to memorise a set of images, or implicit, by simply asking participants to examine a set of images. In the retrieval phase, participants were asked to respond the level of the familiarity as well as the confidence level for their responses for a set of images containing some of the previously-shown images. Pupil size was measured during both the encoding and retrieval phases. In the memorisation phase, smaller pupil size was found for visual stimuli that were subsequently correctly identified as familiar in the retrieval phase. Similarly, in the implicit encoding phase, smaller pupil size was observed for novel stimuli compared to familiar stimuli, although the difference of pupil size between two conditions was marginal. In the retrieval phase, large pupil size was observed for items judged to be familiar compared to items judged to be novel. These findings indicated that memory processes could induce either dilation or contraction of the pupil, based with opposite effects of on memory encoding (contraction) and retrieval (dilation).

In the auditory modality, Papesh, Goldinger, and Hout (2012) used a similar experimental paradigm to study the relationship between pupil size and auditory recognition. They used a mixture of nonwords and words, spoken by either male or female speakers. During the encoding phase, participants simply listened to a set of items, to study presented items for the following test phase. During the test phase, participants heard another set of half new and half old words and nonwords from either the same voice that they heard during the encoding phase or a different voice, to respond whether each word item was "*new*" or "*old*", with three different confidence levels. During the encoding phase, greater pupil dilation was found for nonwords compared to words, and for items that led to correct responses in the subsequent test phase. During the test phase, greater pupil dilation was found for correct responses, the highest confidence level, and items that were spoken by the same voice from the encoding phase, suggesting an effect on pupil size of both memory strength, confidence, and familiarity in the response.

The neural basis of the links between memory and pupil size has been extensively discussed (for a review, see Sara, 2009). Pupil size is of course modulated by exposure, through an interaction of the parasympathetic and sympathetic pathways (Steinhauer, Siegle, Condray, & Pless, 2004). However, at constant light exposure, brain activations related to pupil size changes overlap with regions implicated in memory related behaviours. For example, the locus coeruleus (LC) in the brainstem is one of the principal areas for synthesising norepinephrine that modulates pupil size (Joshi, Li, Kalwani, & Gold, 2016). Concurrently, LC activation is also known to be related to memory encoding

(Lemon et al., 2009; Takeuchi et al., 2016). Furthermore, in the study of Joshi et al. (2016), the inferior colliculus, superior colliculus, anterior and posterior cingulate cortex, which are neural circuits connected to LC, were emphasised as involved pupil size changes. Again, the cingulate cortex is widely known for its involvement in memory processes (Bird et al., 2015; Sutherland, Wishaw, & Kolb, 1988). Thus, there are plausible neural bases for linking pupil dilation and contraction and memory, justifying the use of pupillometry both as a practical index of memory formation but also relating specific neural circuits to memory processes.

One common feature of the previous studies mentioned above was they mostly focused involved explicit and active memory retrieval. Here, we used our auditory paradigm that measures incidental learning using a task that is not an explicit memory retrieval. We used a very similar paradigm to Agus et al. (2010), with a within-trial repetition detection task using random noise sequences. Here, we modified slightly the stimuli (3 within-trials repeats instead of 1) to ensure high behavioural performance on the memory task and tracked pupil dynamics as listeners performed the task.

5.3 Experiment

Methods

Participants

Fourteen volunteers (age = 19 – 36, $M = 25.15$, $SD = 4.86$, 6 male) who had normal or corrected-to-normal vision took part in the experiment. All participants passed an audiometry test showing pure-tone thresholds lower than 20 dB HL for 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. No participants had prior experience on similar experiments. They gave written consents to their participation, and they were paid 10 euros per hour at the end of the experiment. The study was approved by the ethical committee of U. Paris Descartes, France (CERES, IRB: 20154000001072).

Stimuli

Each noise sequence lasted 1.5 s, either of fully random noise (N), or 3 repetitions of a 0.5 s noise segment (repeated noise, RN). Both types of sequences were regenerated afresh on each trial. For each experimental block, one specific RN re-occurred identically over several trials (reference repeated noise, RefRN).

Each of these sequences was embedded in additional noise segments before and after the sequence, to ensure random start and end time of the stimuli and reduce temporal expectations on the pupil signal. The pre-sequence noise had a randomly chosen duration between 0.8 and 1.2 s and the post- sequence noise also lasted between 1.2 s and 0.8 s, with its duration selected such that the whole stimulus duration was always 3.5 s.

After each trial, there was 5 s of silence to let the pupil size return to baseline.

Apparatus

We used the EyeTribe version 1 eyetracker to record eye movements and pupil size during the task. The acquisition sampling frequency was 60 Hz. The audio stimuli were generated and presented with Matlab on a Windows-based laptop, and played through an RME Fireface UC soundcard and Beyer Dynamic DT 770 Pro headphones. Experiments took place in a single-walled sound-attenuated booth (IAC acoustics 400-A). The luminance in the testing booth was kept at 27 cd/m² using DC lamps. Subjects fixated a white fixation cross presented in the middle of a grey background screen (51 cd/m²).

Procedure

Before each block, the eyetracker was calibrated using the Eyetribe calibration procedure. Participants were asked to look at a circle on the screen that moved to five different positions (four corners and the centre of the screen).

Participants then completed 4 test blocks. They were asked to fixate their eyes at a white cross in the middle of the grey screen and minimise blinks especially during the sequence presentation. After each stimulus, they were allowed to provide their response, "yes" or "no" to the task question whether they heard the repetition or not, by means of a computer keyboard. Responses of yes and no were further divided into two different confidence levels. Thus, participants were asked to make responses among 4 choices: "no, very sure", "no", "yes", and "yes, very sure", to consider both a task performance and a confidence level on their responses. After their response, a 5 s silent pause was enforced before next trial to allow the pupil size to return to baseline.

Before the main experiment, participants received a practice session just as in the previous chapters (Agus et al., 2014). It consisted of N and RN stimuli with a high number of within-trial repeats for N, to familiarize participants with the repetition detection task and stimuli. The practice session consisted of three blocks. The first block had 5 s of noise sequences, corresponding to 10 repeats of 0.5 s noise segments for RN.

Subsequent blocks were presented with 4 or 3 repetitions of RN. Auditory feedback was provided during this practice session only.

Data analyses

Due to a relatively low task difficulty, a number of trials between two confidence levels (very sure or not) for "yes" and "no" responses was unbalanced. Therefore, we did not consider conducting an analysis to compare pupil dynamics between confidence levels this time. Responses of different confidence levels were therefore merged into only two categories: "yes" and "no" responses.

Behavioural data

We transformed the "yes" and "no" responses of participants to the sensitivity index d' of signal detection theory. A paired t -test was conducted to compare the d' values between RN and RefRN stimulus types. We also analysed the time course of performance for RefRN and RN, as in previous experiments, to whether performance was constant or changed during a block (Agus et al., 2010; Kang et al., 2017). Hit rates for each trial for both stimulus types were fitted with either a flat line indicating no change on performance (1-parameter model) or an exponential line indicating performance change across trials (3-parameters model), using a least-squared method and penalizing models with extra parameters (Agus et al., 2010; Motulsky & Christopoulos, 2004).

Pupil data

The pupil diameter data were first pre-processed to remove blinks, artefacts, and periods of incorrect fixation. To do so, we took the raw pupil traces recorded during each trial and calculated the mean and standard deviation (SD) of pupil size across trials for each block. We marked missing values outside the range of $\pm 2 SD$ about the mean as eye blinks or artefacts. We also marked values where the gaze direction was further than 2 degrees away from the fixation point as incorrect fixation. Marked values were then removed and replaced by a linear interpolation between values immediately before and after the marked values.

We then verified whether all subjects showed the expected pupil dynamics to sound onset from silence. To do so, we first calculated the baseline pupil size by averaging the pupil size in the 5 s preceding the presentation of each stimulus. We then calculated z -scores of the pupil size during 5 s of before and after each sound onset. All but one of the

subjects showed a clear pupil dilation accompanying the transition from silence to sound. One subject showed the opposite effect of pupil contraction after sound onset and was excluded from further analyses.

A three-point moving average smoothing filter was applied to pre-processed data to remove any high-frequency artefacts. The baseline pupil size was calculated by averaging the pupil size in the 1 s preceding the presentation of each sequence. Baseline was then subtracted to the pupil traces during the sequence presentation. To observe the effects of learning, we compared both peak pupil dilation values and pupil traces between the first 5 and last 5 presentation of each stimulus type. We computed a peak of pupil dilation by choosing a time point of the maximum pupil value and averaging pupil values before and after 0.5 s of that time point for each trial and stimulus type. A repeated-measures ANOVA was conducted to compare peak pupil values between stimulus types (RN and RefRN) and trial number (first 5 and last 5). For pupil traces, we averaged the data for each participant for the first 5 presentation of each stimulus type on the one hand, and for the last 5 presentations of each stimulus type on the other hand. A cluster permutation test (Maris & Oostenveld, 2007) was applied to test for the statistical significance of the effects across the group of participants.

5.4 Results

Behavioural results

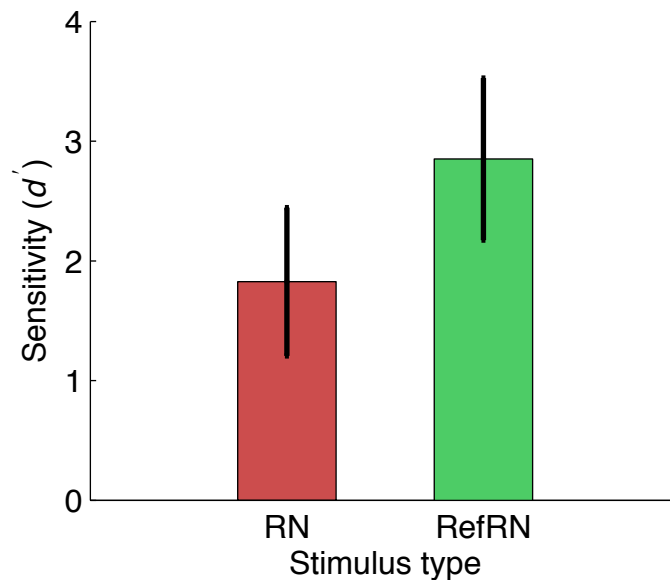


Figure 5-1. Average performance measured by the sensitivity index d' for RN (red) and RefRN (green). Error bars are 95% of confidence intervals.

Figure 5-1 displays the average d' observed for the behavioural task across the group of participants. As expected, performance was relatively high because of the larger number of within-trial repeats, and most importantly performance was higher for RefRN than RN. A paired t -test confirmed the effect of stimulus type on d' ($t(13) = -6.55, p < 0.001$). Most of the experimental blocks showed high performance. In particular, testing for the criterion used to defined “good blocks” in previous chapters, we observed that only 10 blocks showed hit rates lower than 90% for the last 10 trials of the block. We therefore included all blocks in the subsequent analyses, without distinguishing between good and bad blocks.

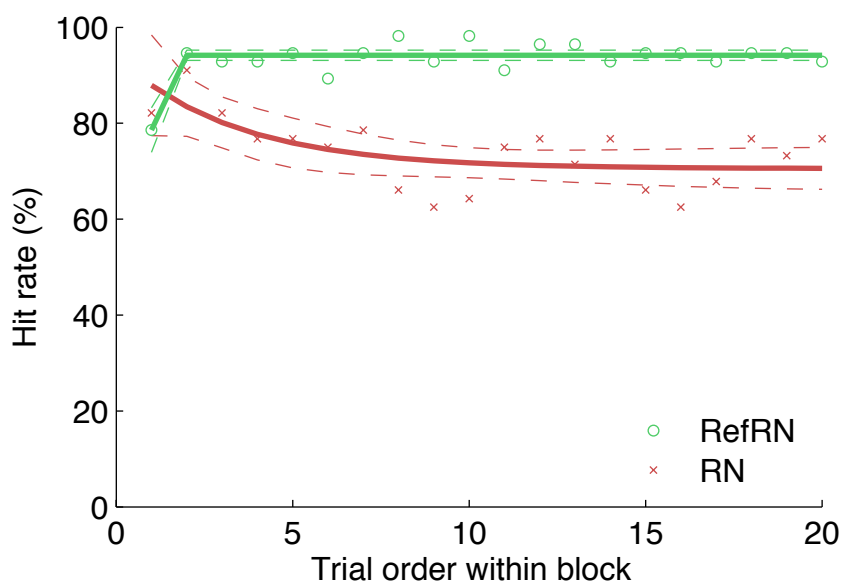


Figure 5-2. Hit rate along trial orders for RN (red crosses) and RefRN (green circles). Dashed lines indicate 95% confidence intervals.

Figure 5-2 displays the performance for each stimulus type according to the trial order during the block. Consistent with the average d' data, performance was generally high even from the first trial for each stimulus type. Nevertheless, we observed a significant performance increase for RefRN ($F(2,17) = 22.26, p < 0.001$). A gradual decrease of performance for RN was also observed ($F(2,17) = 6.60, p < 0.01$), likely due to the criterion effect (Agus et al., 2010). Even if RefRNs were at ceiling for most of the block, it is still possible that they became easier and easier to detect, causing the gradual shift in criterion.

Pupil dynamics

Figure 5-3 displays the pupil size traces during each stimulus sequence, averaged for the first 5 trials or the last 5 trials of each stimulus type. For the first 5 trials, a similar gradual pupil dilation was observed for all stimulus types, with peak dilation obtained at 3 – 4 s after the onset of the sequence. For the last 5 trials, there was no visible change in pupil size compared to the first 5 trials in the RN condition, a small decrease in the N condition, and a larger decrease in the RefRN condition.

A repeated-measures ANOVA showed that the peak of pupil size did not overall differ across stimulus type ($F(2,24) = 1.60, p > 0.05$). The effect of position during the block (first 5 trial or last 5 trials) had a marginal overall effect on pupil dynamics ($F(1,23)$

= 4.51, $p = 0.055$). No significant interaction between the stimulus type and the effect of trial number was observed ($p > 0.05$).

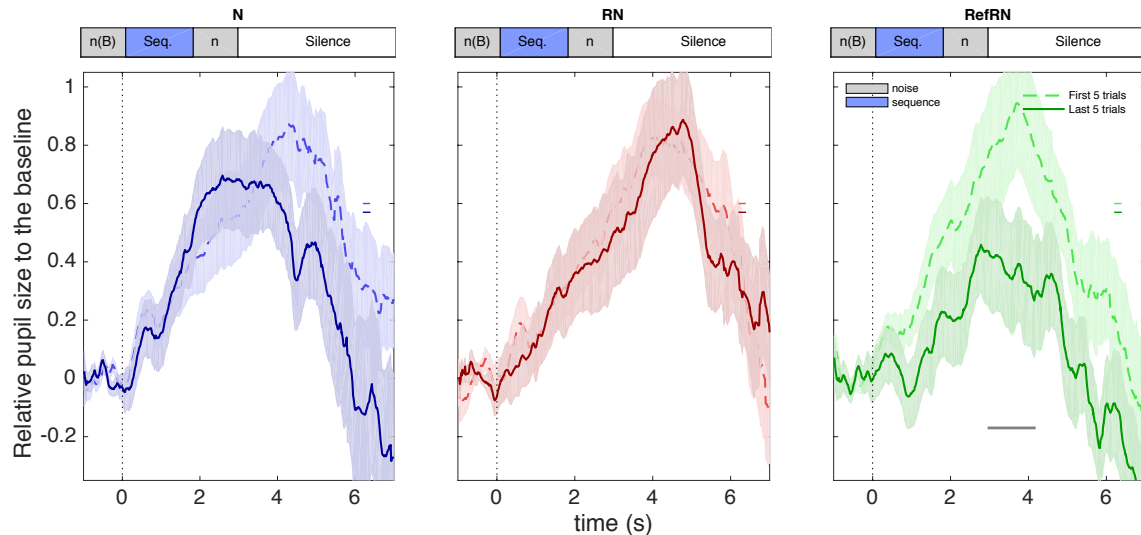


Figure 5-3. Pupil dynamics on average of first 5 trials (dashed lines) and last 5 trials (solid lines) for each condition (N: blue, RN: red, and RefRN: green). A bar panel at the top of each graph indicates the time bin of pre- and post-noise (grey), stimuli (cobalt blue), and silence (white). Response time was always around 6 s after the stimulus onset (marked as colour-coded short horizontal lines for each condition and trial order). A dark grey horizontal line at the bottom of the RefRN graph indicates the time period where $p < 0.05$, based on the cluster permutation test. Error shades indicate 95% confidence intervals.

To focus on the possible effect of learning on pupil size, we directly compared pupil size between the first 5 trials and the last 5 trials, for each stimulus type separately. To do so, we ran a cluster permutation. For RefRN, the permutation test showed that the pupil size was significantly decreased at the end of the experiment, and this difference appeared between 2.95 s and 4.18 s after sequence onset ($p < 0.05$; between 3.32 – 3.9 s: $p < 0.01$). No such differences were observed at any time for the N and RN stimulus types.

5.5 Discussion

The present study recorded pupil size changes during the memory for noise experimental paradigm. The behavioural results confirmed yet again a rapid formation of auditory memory traces, as found in previous studies found (e.g., Agus et al., 20210; Andrillon et

al., 2015). Here we used a relatively easy task with 3 within-trial repeats, to ensure that most participants (naïve) would display learning on most blocks, as the main interest was the pupil measurement. Even so, all hallmarks of perceptual learning were observed: an advantage in performance for RefRN over RN, and changes in performance along the block consistent with learning.

The pupil results demonstrated a significant decrease of pupil size only for RefRN between the beginning and the end of an experimental block. Interestingly, only RefRN could be learnt over a block, as it re-occurred several times and saw an increase in performance. For N and RN trials, the nature of the task did not change over a block as these sound sequences were novel on each trial. Consistent with this observation, there was no significant change in pupil size for these stimulus conditions, even though a trend for a decrease in pupil dilation may have been present for N sequences. These findings suggest that the pupil signal may be correlated to task performance in the memory for noise paradigm. Here, we did not test for quantitative correlations on a participant per participant basis, as performance was high overall.

The interpretation of these pupil findings, still preliminary, could be as follow. The pupil dilation at the beginning of a block probably simply reflected the salient change from silence to sound presentation and the accompanying need to provide a perceptual decision for participants (de Gee, Knapen, & Donner, 2014).

The specific effect on RefRN after learning of RefRN may be related to memory processes. We observed a smaller pupil dilation for RefRN compared to N and RN, which is similar to the findings of Kafkas and Montaldi (2011) who reported a gradual decrease of pupil dilation for visual stimuli that were recognised with the greater familiarity level. However, the effect seems opposite to that of Papesh et al (2012). The encoding phase for the study of Papesh et al. involved intentional memorisation, unlike the study of Kafkas and Montaldi as well as the present study that more focused on incidental learning. Thus, this difference could have been caused by the different encoding phase used in the study of Kafkas and Montaldi. A novel finding here is that the effect on pupil size was observed even without asking participants for an explicit memory retrieval task. Participants were not told of re-occurring RefRN, and the task did not involve explicitly recognising RefRN, simply within-trial repetition detection just as in the RN trials.

There is however an alternative explanation to the pupil results. The performance for RefRN improved after learning. Also, the task may become very different before learning, where within-trial comparisons are needed, compared to after learning, where

recognition of a single learnt segment is sufficient to provide an answer (Agus & Pressnitzer, 2013). Reaction times are also known to shorten after learning (Andrillon et al., 2015). It could thus be that the task became easier and required less attention or effort for RefRN after learning. There are several reports linking listening effort or confidence in the response to pupil size (Kafkas & Montaldi, 2011; Koelewijn et al., 2015; Lempert, Chen, & Fleming, 2015; van Rijn, Dalenberg, Borst, & Sprenger, 2012). As less effort and more confidence are generally related to smaller pupil dilation, these factors could either fully or partly explain the pupil results observed here.

In future experiments that have already been initiated, we will record pupil dynamics using similar stimuli but without an explicit task. A similar approach was used in Andrillon et al., 2015 (supplementary experiment), who observed EEG correlates of learning even as participant's attention was focused away from the experimental stimuli of interests. Moreover, behavioural and EEG evidence of learning can be observed during sleep (Andrillon et al., 2017). Using these new conditions without a memory-related task, we should be able to parcel out task-related and memory-related influences on pupil dilation.

Chapter 6

Conclusions

The thesis presented a series of experimental studies aimed at exploring memory processes for temporal information in sensory systems. We tested for the possibility of rapid learning of temporal patterns in audition with normal hearing listeners, in touch and in vision with matched experimental paradigms, and in cochlear implant users. The novel findings provided the first evidence of rapid learning of temporal patterns in audition, and suggested the existence of common mechanisms of rapid incidental learning across modalities. We also explored a new experimental tool, pupillometry, for the study of rapid memory formation in audition. This Conclusion chapter will first summarise the main findings of each study and then present perspectives for future extensions of the present work.

6.1 Main findings

Study 1. Psychoacoustics in NH listeners:

Can NH listeners show fast and robust auditory learning for sounds containing only temporal information?

We first studied NH listeners' and tested their ability to learn sounds that contained only timing information. To devise such sounds, we used click trains with random inter-click intervals (ICIs) drawn from either uniform or Poisson distributions. Following the paradigm we used throughout the thesis, the click trains contained either a 2-s long sequence of random intervals (random click trains, C) or a 1-s long sequence of random intervals that was immediately repeated within a trial (repeated click trains, RC). Random time intervals were generated afresh for both stimulus types for each trial. For a last type of trial, one “frozen” RC re-occurred in multiple, randomly positioned trials within a

CONCLUSIONS

session (reference repeated click trains, RefRC). Without being told about the existence of RefRCs, listeners were asked to report the presence or absence of a within-trial repetition for each click train. Listeners showed superior performance for RefRC compared to RC overall, and we observed a performance improvement along trials for RefRC. This performance difference between RefRC and RC combined to the improvement for RefRC over time were regarded as evidence that listeners learnt the time pattern that re-occurred (RefRC). Learning was observed only after a few appearances of RefRC, and the improved performance was robust throughout the session once it had been learned. The general task performance was better when click trains were produced with ICIs drawn from the Poisson distribution, which was associated to larger variance across ICIs compared to uniformly distributed ICIs.

We also demonstrated that the memory for RefRC was sensitive to the sequential order of intervals. When the order of the time intervals of the learned temporal pattern was reversed (reversed RefRC), listeners did not retain their performance on the within-trial repetition detection task, unlike what was observed for Gaussian white noise in the study of Agus et al. (2010). Rather, results were as if listeners treated the reversed RefRC as a new random time pattern. This suggests that recognition and learning of RefRC did not rely on local cues, such as time intervals especially long or short that may have been especially salient and easier to memorize. Rather, the information encoded in memory included the sequential order of time intervals.

Study 2. Multimodality comparisons for memory of time patterns:

Can sensory modalities benefit from similar mechanisms to memorise temporal information? If so, can learning in one modality transfer to another?

Meaningful temporal information exists in all sensory systems, but whether the processing of such information follows common mechanisms or independent mechanisms across modalities is still a subject of investigation. Based on our findings of rapid learning of time patterns in audition in the first study, we examined the learning of time patterns in two other sensory modalities: touch and vision.

Experiment 1 in the second study (Chapter 3) investigated temporal pattern processing within each modality of audition, touch, and vision. We carefully designed the stimuli and procedure to determine a set of common parameters for time patterns in all modalities, with the constraints that individual time intervals should be perceived in all

three modalities. We then examined the process of temporal pattern learning in each modality using the same procedure across modalities, again the repetition detection task. General performance was best in audition, intermediate in touch and worst in vision. This parallels the quantitative difference in temporal sensitivity across modalities previously reported in several types of discrimination tasks. However, regardless of such difference, we observed a similar effect of learning in all three modalities, in the form of an advantage of the reference time patterns and a rapid improvement in performance over re-occurrences. This suggests similar mechanisms for learning of temporal information in different sensory modalities.

We then performed another experiment to examine a possible carry-over of memory between modalities. Keeping the same experimental parameters, we chose audition as the initial learning modality since it showed the strongest learning effects. After using the standard repetition-detection task to trigger the learning of RefRPs in the auditory modality, we switched the presentation mode to either touch or vision, and continued the same repetition-detection task either with the same RefRP (Transfer condition) or a new RefRP (Control condition). Initial performance for RefRP in the Transfer condition was significantly better than in the Control condition, in both touch and vision. Again, general performance in vision was slightly worse than in touch but the transfer effect was as strong for both modalities such that there was no interaction between modality and condition. The transfer effect found for both touch and vision from audition suggests that either the learning of temporal information recruits common mechanisms across modalities, or that there are strong and relatively early interactions across modalities for such mechanisms.

Study 3. Psychoacoustics with CI listeners:

Can temporal patterns of electrical pulses in a cochlear implant be learnt by CI listeners as efficiently as similar patterns presented acoustically to NH listeners?

We expanded our investigation of rapid learning of time patterns to CI listeners, as such listeners are known to have access to peripheral cues representing temporal information, but have not been probed yet on their ability to learn such cues. We applied the same experimental paradigm to test for rapid learning, this time using a series of pulses directly transmitted to the electrodes of the CI device. First, to study the learning of low-rate temporal patterns, we used pulse trains with average pulse rate of 10 Hz, 40 Hz,

CONCLUSIONS

simultaneously transmitted on 4 equally spaced electrodes. Furthermore, in CI listeners only, we presented a fast-rate sequence at 300 Hz, which would have been difficult to present without introducing spectral cues with acoustical stimulation. CI listeners showed fast and clear learning of RefRP in all rate conditions, despite a trend of decreasing performance with increasing rates. General performance of CI listeners was slightly lower than for NH listeners. This could have resulted from mean age differences between two groups, as a correlation between performance and age was observed (NH mean age: 22.5, CI mean age: 55). Regardless of this difference, the performance improvement resulting from the learning was similar between the two listening groups, indicating the preserved ability of CI listeners to rapidly learn auditory information in the time domain.

Study 4. Pupil size change during auditory learning:

Can changes in pupil size be used as an index of auditory memory formation? If so, what is the link between pupil size and memory?

Pupillometry is a physiological measurement increasingly used to study perceptual processing, which is also technically simpler compared to other neuro-imaging tools such as EEG, MEG, or fMRI. As a final experimental study of the thesis, we measured pupil size during the repetition detection task for random noise, to test whether changes in pupil size would be indicative of auditory memory. We used Gaussian noise (denoted as N, RN, and RefRN which correspond to C, RC, and RefRC respectively), instead of click trains, as there are past investigations of neural correlates of such noise stimuli. The amount of pupil dilation in response to the noise sequences for early trials did not differ across stimulus types. However, the pupil size became significantly smaller during the last few presentations for RefRN while no change in pupil size was observed for either RN or N. The large effect of the task on pupil size could be a correlate of memory processes, but they could also at least partly be due to the attentional load required during performing the task, as the task for RefRN trials becomes easier after learning, compared to a constant difficulty for RN or N trials. Overall, the result suggests that pupillometry can be used to index auditory memory formation during an active task, and further experiments are required to determine the cause of the pupil effect.

6.2 Perspectives

The results of experimental studies in the thesis, in my opinion, have been successful to provide new insights on temporal processing in sensory systems. The task paradigm we used can be effectively applied in both psychophysics for different modality and for neuro-physiological investigations of sensory memory. Further investigations on the topic would be useful to further clarify and understand the stimulus characteristics impacting performance on our task and the underlying neural mechanisms.

From the findings of Chapter 2, we proposed that the sequential order of temporal patterns was important for time pattern recognition. To support this claim, a complete time reversal was applied to a time pattern after it had been learnt. Future experiments could test whether listeners would be tolerant of more subtle perturbations of the learned time pattern. Temporal changes such as introducing a jitter only to one (or a few) of the intervals could be imposed to the learned patterns, and we could observe listeners' ability to recognise the sequences with jitter. This investigation would enable us to track down how sensitive we would be to time changes for temporal pattern recognition and how resistant we would be to these time changes.

For the study of temporal learning in different sensory modalities in Chapter 3, we chose to use the exact same stimulus and task parameters across all modalities, to produce time intervals above the thresholds for perceptual fusion across modalities. This was to allow for a fair comparison across modalities, considering the quantitative differences in temporal sensitivity that are known to exist across modalities. However, difference in baseline performance was still observed across modalities. The best performance was observed when temporal patterns were presented in the auditory modality, followed by tactile and visual modalities, which follows the same order as the temporal acuity in those modalities. Another way to design the experiments would be to match task difficulty across modalities, for instance by adjusting the mean ICIs to have the same performance on the within-trial repetition detection task. In this case, the amount of learning could be compared across modalities by simply considering performance for the reference patterns.

A follow-up study on the relationship between pupil dynamics and auditory information processing is finally needed. To disentangle the effects of memory and task demands, pupil size changes should be recorded in a passive listening condition where it can still be reasonably hypothesised that incidental learning occurs. The comparison of

CONCLUSIONS

results with and without an explicit task would help parcel out the role of memory formation and task demands on changes in pupil size.

In conclusion, the experimental work presented in my thesis attempted to focus on mechanisms of memory in the time domain across sensory systems, by adapting a paradigm previously used to study the learning of auditory information. The thesis revealed a remarkable ability to learning time patterns for audition in normal hearing listeners, but also for touch, vision, and electric hearing in cochlear implant listeners. The first results concerning pupil dynamics further suggest new tools to investigate the underlying mechanism of such learning. In an appendix to the thesis, I also present a short report of ongoing investigations using electrophysiological recordings in an animal model. The past results and ongoing investigations should help us understand how we seem to be able to memorise complex sensory information without much effort, rapidly, which is surely useful to analyse complex sensory scenes.

References

- Abel, S. M. (1972). Discrimination of Temporal Gaps. *The Journal of the Acoustical Society of America*, 52(2B), 519–524. <http://doi.org/10.1121/1.1913139>
- Agus, T. R., Carrión-Castillo, A., Pressnitzer, D., & Ramus, F. (2014). Perceptual Learning of Acoustic Noise by Individuals With Dyslexia. *Journal of Speech Language and Hearing Research*, 57(3), 1069–1077. [http://doi.org/10.1044/1092-4388\(2013/13-0020\)](http://doi.org/10.1044/1092-4388(2013/13-0020))
- Agus, T. R., Jeannou, C., & Pressnitzer, D. (2013). Individual difference for pitched features learnt in white noise by musical listeners. Presented at the British Society of Audiology annual conference, Newcastle-under-Lyme, UK.
- Agus, T. R., & Pressnitzer, D. (2013). The detection of repetitions in noise before and after perceptual learning. *The Journal of the Acoustical Society of America*, 134, 464–473. doi:10.1121/1.4807641
- Agus, T. R., Suied, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America*, 131, 4124–4133. doi:10.1121/1.3701865
- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid Formation of Robust Auditory Memories: Insights from Noise. *Neuron*, 66(4), 610–618. <http://doi.org/10.1016/j.neuron.2010.04.014>
- Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, 26(5), 340–354. <http://doi.org/10.3758/BF03204158>
- Allan, L. G. (1998). The influence of the scalar timing model on human timing research. *Behavioural Processes*, 44, 101–117.
- Allen, D., Hees, R. F., & Nordby, K. (1998). Is the rod visual field temporally homogeneous? *Vision Research*, 38, 3927–3931.
- Andreou, L. V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing research*, 280(1), 228-235.
- Andrillon, T., Kouider, S., Agus, T., & Pressnitzer, D. (2015). Perceptual Learning of Acoustic Noise Generates Memory-Evoked Potentials. *Current Biology*, 25(21), 2823–2829. <http://doi.org/10.1016/j.cub.2015.09.027>
- Andrillon, T., Pressnitzer, D., Léger, D., & Kouider, S. (2017). Formation and

- suppression of acoustic memories during human sleep. *Nature Communications*, 8, 179. doi:10.1038/s41467-017-00071-z
- Bale, M. R., Bitzidou, M., Pitas, A., Brebner, L. S., Khazim, L., Anagnou, S. T., et al. (2017). Learning and recognition of tactile temporal sequences by mice and humans. *eLife*, 6, e27333. <http://doi.org/10.7554/eLife.27333>
- Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, 25(2), R60–R61. <http://doi.org/10.1016/j.cub.2014.12.011>
- Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, 41(3), 254–311. <http://doi.org/10.1006/cogp.2000.0738>
- Bendor, D., & Wang, X. (2007). Differential neural coding of acoustic flutter within primate auditory cortex. *Nat Neurosci*, 10, 763–771. doi:10.1038/nn1888
- Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the Human Neocortex to the Temporal Dynamics of Attended Events. *Journal of Neuroscience*, 31(9), 3176–3185. <http://doi.org/10.1523/JNEUROSCI.4518-10.2011>
- Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Attention, Perception, & Psychophysics*, 64(5), 844-854.
- Bidelman, G. M., Schug, J. M., Jennings, S. G., & Bhagat, S. P. (2014). Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. *The Journal of the Acoustical Society of America*, 136, EL33–EL39. doi:10.1121/1.4885484
- Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J., & Burgess, N. (2015). Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *Journal of Neuroscience*, 35(43), 14426-14434.
- Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat Processing Is Pre-Attentive for Metrically Simple Rhythms with Clear Accents: An ERP Study. *PLoS ONE*, 9(5), e97467–9. <http://doi.org/10.1371/journal.pone.0097467>
- Bratzke, D., Seifried, T., & Ulrich, R. (2012). Perceptual learning in temporal discrimination: asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research*, 221(2), 205–210. <http://doi.org/10.1007/s00221-012-3162-0>
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- Bueti, D., Lasaponara, S., Cercignani, M., & Macaluso, E. (2012). Learning about Time:

REFERENCES

- Plastic Changes and Interindividual Brain Differences. *Neuron*, 75(4), 725–737.
<http://doi.org/10.1016/j.neuron.2012.07.019>
- Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, 198(1), 49–57.
<http://doi.org/10.1007/s00221-009-1933-z>
- Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci*, 22, 2886–2902.
- Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones I Pitch and pitch salience. *Journal of Neurophysiology*, 76, 1698–1716.
- Chen, L., & Zhou, X. (2014). Fast transfer of crossmodal time interval training. *Exp Brain Res* 232, 1855–1864.
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*, 25(5), 975–979.
<http://doi.org/10.1121/1.1907229>
- Church, R. M., & Gibbon, J. (1982). Temporal generalization. *Journal of Experimental Psychology: Animal Behavior Processes*, 8(2), 165.
- Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition*, 28, 529–538.
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychological Science*, 17, 905–912.
- Craig, J. C. (1973). A constant error in the perception of brief temporal intervals. *Perception & Psychophysics*, 13(1), 99–104. <http://doi.org/10.3758/BF03207241>
- Creelman, C. D. (1962). Human Discrimination of Auditory Duration. *The Journal of the Acoustical Society of America*, 34(5), 582–593. <http://doi.org/10.1121/1.1918172>
- de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, 111(5), E618–E625. <http://doi.org/10.1073/pnas.1317557111>
- Demany, L., & Semal, C. (2007). *The Role of Memory in Auditory Perception*. In Auditory Perception of Sound Sources, Springer Handbook of Auditory Research, Springer US, Boston, MA, Vol. 29, pp. 77–113. doi:10.1007/978-0-387-71305-2_4
- Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54(3), 277–286.

- <http://doi.org/10.3758/BF03205262>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2), 179-188.
- Fraisse, P. (1948). Fraisse, P. (1948). II.-Rythmes auditifs et rythmes visuels. *L'année psychologique*, 49(1), 21-42.
- Fujisaki, W., & Nishida, S. (2009). Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception. *Experimental Brain Research*, 198(2-3), 245-259. <http://doi.org/10.1007/s00221-009-1870-x>
- Gaudrain, E., Deeks, J. M., & Carlyon, R. P. (2017). Temporal Regularity Detection and Rate Discrimination in Cochlear-Implant Listeners. *Journal of the Association for Research in Otolaryngology*, 18(2), 387-397. <http://doi.org/10.1007/s10162-016-0586-4>
- Gault, R. H., & Goodfellow, L. D. (1938). An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. *Journal of General Psychology*, 18(1), 41-47. <http://doi.org/10.1080/00221309.1938.9709888>
- Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience*, 17, 732-737.
- Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *J Exp Psychol*, 71, 378-381.
- Gfeller, K., & Lansing, C. R. (1991). Melodic, Rhythmic, and Timbral Perception of Adult Cochlear Implant. *Journal of Speech and Hearing Research*, 34, 916-920.
- Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., & Knutson, J. F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and Hearing*, 18(3), 252-260. <http://doi.org/10.1097/00003446-199706000-00008>
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279-325. <http://doi.org/10.1037//0033-295x.84.3.279>
- Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York Academy of Sciences*, 423, 52-77.
- Glenberg, A. M., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, 19(5), 514-522. <http://doi.org/10.3758/BF03199576>

REFERENCES

- Glenberg, A. M., Mann, S., Altman, L., Forman, T., & Prociase, S. (1989). Modality effects in the coding and reproduction of rhythms. *Memory & Cognition*, *17*(4), 373–383.
- Gold, J. M., Aizenman, A., Bond, S. M., & Sekuler, R. (2014). Memory and incidental learning for visual frozen noise sequences. *Vision Res*, *99*, 19–36.
- Goldinger, S. D., & Papesh, M. H. (2012). Pupil Dilation Reflects the Creation and Retrieval of Memories. *Current Directions in Psychological Science*, *21*(2), 90–95. <http://doi.org/10.1177/0963721412436811>
- Goldstone, S., & Goldfarb, J. L. (1963). Judgment of filled and unfilled durations: Intersensory factors. *Perceptual and Motor Skills*, *17*, 763–774.
- Goodfellow, L. D. (1934). An Empirical Comparison of Audition, Vision, and Touch in the Discrimination of Short Intervals of Time. *The American Journal of Psychology*, *46*(2), 243–258.
- Gorea, A. (2015). A Refresher of the Original Bloch's Law Paper (Bloch, July 1885). *i-Perception*, *6*, 204166951559304–204166951559306.
- Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Exp Brain Res* *220*, 51–61.
- Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). fMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. *NeuroImage*, *54*(2), 1231–1243. <http://doi.org/10.1016/j.neuroimage.2010.09.033>
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *The Journal of the Acoustical Society of America*, *108*(1), 263–271.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, *54*(3), 383–394.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, *72*(3), 561–582. <http://doi.org/10.3758/APP.72.3.561>
- Grondin, S. (2014). Why studying intermodal duration discrimination matters. *Frontiers in psychology*, *5*, 628. <http://doi.org/10.3389/fpsyg.2014.00628>.
- Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*(3), 245–256.
- Gur, M., & Snodderly, D. M. (1997). A dissociation between brain activity and

- perception: chromatically opponent cortical neurons signal chromatic flicker that is not perceived. *Vision Res*, 37, 377–382.
- Guttman, N., & Julesz, B. (1963). Lower Limits of Auditory Periodicity Analysis. *The Journal of the Acoustical Society of America*, 35(4), 610–610. <http://doi.org/10.1121/1.1918551>
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science*, 16(3), 228–235. <http://doi.org/10.1111/j.0956-7976.2005.00808.x>
- Handel, S., & Buffardi, L. (1968). Pattern perception: integrating information presented in two modalities. *Science*, 162, 1026–1028.
- Hansen, L., & Cottrell, D. (2013). An Evaluation of Modality Preference Using a “Morse Code” Recall Task. *The Journal of Experimental Education*, 81, 123–137.
- Hardy, N.F., & Buonomano, D. V. (2016). Neurocomputational models of interval and pattern timing. *Current Opinion in Behavioral Sciences*, 8, 250–257.
- Harris, J. A., Harris, I. M., & Diamond, M. E. (2001). The topography of tactile learning in humans. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, 21(3), 1056–1061.
- Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19, 304–313. doi:10.1016/j.tics.2015.04.006
- Hirsh, I. J. (1959). Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America*, 31(6), 759–767. <http://doi.org/10.1121/1.1907782>
- Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, 47(3), 215–226.
- Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, 12(7), 273–280. <http://doi.org/10.1016/j.tics.2008.04.002>
- Jones, M. R., & Boltz, M. (1989). Dynamic Attending and Responses to Time. *Psychological Review*, 96(3), 459–491.
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron*, 89(1), 221–234. <http://doi.org/10.1016/j.neuron.2015.11.028>
- Kaernbach, C. (1993). Temporal and spectral basis of the features perceived in repeated noise. *The Journal of the Acoustical Society of America*, 94, 91–97.

REFERENCES

- Kaernbach, C. (1999). A behavioral reverse correlation technique to decipher early auditory feature coding. *The Journal of the Acoustical Society of America*, *105*, 966–966. doi:10.1121/1.425294
- Kaernbach, C. (2000). *Early auditory feature coding*, In A. Schick, M. Meister, and C. Reckhardt (Eds.), *Contributions to psychological acoustics: Results of the 8th Oldenburg Symposium on Psychological Acoustics*, Oldenburg, pp. 295–307.
- Kaernbach, C. (2004). The Memory of Noise. *Experimental Psychology*, *51*(4), 240–248. <http://doi.org/10.1027/1618-3169.51.4.240>
- Kaernbach, C., & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *The Journal of the Acoustical Society of America*, *104*, 2298–2306.
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *The Quarterly Journal of Experimental Psychology*, *64*(10), 1971–1989. <http://doi.org/10.1080/17470218.2011.588335>
- Kalloniatis, M., & Luu, C. (1995). *Temporal Resolution*. In: *Webvision The Organization of the Retina and Visual System* (Kolb H, Fernandez E, Nelson R, eds). Salt Lake.
- Kanai, R., Lloyd, H., Bueti, D., & Walsh, V. (2011). Modality-independent role of the primary auditory cortex in time estimation. *Experimental Brain Research*, *209*(3), 465–471. <http://doi.org/10.1007/s00221-011-2577-3>
- Kang, H., Agus, T. R., & Pressnitzer, D. (2017). Auditory memory for random time patterns. *The Journal of the Acoustical Society of America*, *142*, 2219. <http://dx.doi.org/10.1121/1.5007730>
- Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, *10*(2), 141–147. <http://doi.org/10.1101/lm.55503>
- Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron*, *53*(3), 427–438. <http://doi.org/10.1016/j.neuron.2007.01.006>
- Klampfl, S., David, S.V., Yin, P., Shamma, S.A., & Maass, W. (2012). A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *Journal of Neurophysiology*, *108*, 1366–1380.
- Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to

- focus attention. *Hearing Research*, 323(C), 81–90.
<http://doi.org/10.1016/j.heares.2015.02.004>
- Kong, Y.-Y., Cruz, R., Jones, J. A., & Zeng, F.-G. (2004). Music Perception with Temporal Cues in Acoustic and Electric Hearing. *Ear and Hearing*, 25(2), 173–185.
<http://doi.org/10.1097/01.AUD.0000120365.97792.2F>
- Krumbholz, K., Patterson, R. D., & Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *The Journal of the Acoustical Society of America*, 108, 1170–1180.
- Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., & Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings of the Royal Society B: Biological Sciences*, 281(1791), 20141000. <http://doi.org/10.1098/rspb.2014.1000>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol*, 4, 863. doi:10.3389/fpsyg.2013.00863
- Lapid, E., Ulrich, R., & Rammsayer, T. (2008). On estimating the difference limen in duration discrimination tasks: A comparison of the 2AFC and the reminder task. *Perception & Psychophysics*, 70(2), 291–305. <http://doi.org/10.3758/PP.70.2.291>
- Large, E. W., & Jones, M. R. (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review*, 106(1), 119–159.
- Lechelt, E. C. (1975). Temporal numerosity discrimination: intermodal comparisons revisited. *British Journal of Psychology*, 66(1), 101–108.
<http://doi.org/10.1111/j.2044-8295.1975.tb01444.x>
- Lemon, N., Aydin-Abidin, S., Funke, K., & Manahan-Vaughan, D. (2009). Locus coeruleus activation facilitates memory encoding and induces hippocampal LTD that depends on β -adrenergic receptor activation. *Cerebral cortex*, 19(12), 2827–2837.
- Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. *PLoS ONE*, 10(5), e0126588–12. <http://doi.org/10.1371/journal.pone.0126588>
- Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, 5, 356–363.
- Lewis, M. F. (1967). Two-Flash Thresholds as a Function of Luminance in the Dark-Adapted Eye. *J Opt Soc Am*, 57, 814–815.
- Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and

REFERENCES

- vibrotactile information for judgments of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1005–1019. <http://doi.org/10.1037/a0015021>
- Lim, Y., Lagoy, R., Shinn-Cunningham, B. G., & Gardner, T. J. (2016). Transformation of temporal sequences in the zebra finch auditory system. *eLife*, 5, e18205. <http://doi.org/10.7554/eLife.18205>
- Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, 4(11), 1131–1138. <http://doi.org/10.1038/mn737>
- Luo, H., Tian, X., Song, K., Zhou, K., & Poeppel, D. (2013). Neural Response Phase Tracks How Listeners Learn New Acoustic Representations. *Current Biology*, 23(11), 968–974. <http://doi.org/10.1016/j.cub.2013.04.031>
- Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. *Current Biology*, 24(18), R878–R884. <http://doi.org/10.1016/j.cub.2014.06.053>
- Macmillan, N. A., & Creelman, C. D. (2004) Detection Theory. Psychology Press.
- Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, 23(11), 1369–1386. <http://doi.org/10.1068/p231369>
- Malapani, C., & Fairhurst, S. (2002). Scalar Timing in Animals and Humans. *Learning and Motivation*, 33(1), 156–176. <http://doi.org/10.1006/lmot.2001.1105>
- Manning, S. K., Pasquali, P. E., & Smith, C. A. (1975). Effects of visual and tactual stimulus presentation on learning two-choice patterned and semirandom sequences. *Journal of Experimental Psychology: Human Learning and Memory* 1(6), 736–744.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, 164(1), 177-190.
- Marks, L. E. (1987). On cross-modal similarity: Perceiving temporal patterns by hearing, touch, and vision. *Perception & Psychophysics*, 42(3), 250–256. <http://doi.org/10.3758/BF03203076>
- Masquelier, T., Guyonneau, R., & Thorpe, S. J. (2008). Spike Timing Dependent Plasticity Finds the Start of Repeating Patterns in Continuous Spike Trains. *PLoS ONE*, 3(1), e1377–9. <http://doi.org/10.1371/journal.pone.0001377>
- McAuley, J. D., & Kidd, G. R. (1998). Effect of deviations from temporal expectations on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 24(6), 1786–1800.

- McDermott, J. H., Wroblewski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences*, *108*(3), 1188-1193.
- Merchant, H., Zarco, W., & Prado, L. (2008). Do We Have a Common Mechanism for Measuring Time in the Hundreds of Millisecond Range? Evidence From Multiple-Interval Timing Tasks. *Journal of Neurophysiology*, *99*, 939–949.
- Moore, B. C., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In *Pitch* (pp. 234-277). Springer New York.
- Moore, B. C., & Gockel, H. E. (2012). Properties of Auditory Stream Formation. *Phil. Trans. R. Soc. B*, *367*(1591), 919-931.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163. [http://doi.org/10.1016/S0926-6410\(03\)00089-2](http://doi.org/10.1016/S0926-6410(03)00089-2)
- Motulsky, H., & Christopoulos, A. (2004). *Fitting Models to Biological Data using Linear and Nonlinear Regression*, Oxford University Press, New York.
- Naatanen, R., Syssoeva, O., & Takegata, R. (2004). Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. *Psychophysiology*, *41*(4), 660–663. <http://doi.org/10.1111/j.1469-8986.2004.00182.x>
- Naber, M., Frassle, S., Rutishauser, U., & Einhauser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11–11. <http://doi.org/10.1167/13.2.11>
- Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-Related Improvements in Somatosensory Interval Discrimination Are Temporally Specific But Generalize across Skin Location, Hemisphere, and Modality. *Journal of Neuroscience*, *18*(4), 1559–1570.
- Nazzaro, J. R., & Nazzaro, J. N. (1970). Auditory versus visual learning of temporal patterns. *J Exp Psychol*, *84*, 477–478.
- Nikolić, D., Häusler, S., Singer, W., & Maass, W. (2009). Distributed Fading Memory for Stimulus Properties in the Primary Visual Cortex. *PLoS Biology*, *7*, e1000260–19.
- Nilsson, T. H. (1979). Visual Temporal Discriminations of Brief Pulse Intervals. *Vision Research*, *19*, 41–48.
- Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal*

REFERENCES

- Society B: Biological Sciences*, 369(1658), 20130393.
<http://doi.org/10.1098/rstb.2013.0393>
- O'Regan, J. K. (2011). *Why Red Doesn't Sound Like a Bell: Understanding the feel of consciousness*. Oxford University Press.
- Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in temporal perception. *Psychonomic Bulletin & Review*, 18, 429–454.
- Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, 48(10), 1346–1353.
<http://doi.org/10.1111/j.1469-8986.2011.01217.x>
- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83(1), 56–64.
<http://doi.org/10.1016/j.ijpsycho.2011.10.002>
- Pasinski, A. C., McAuley, J. D., & Snyder, J. S. (2015). How modality specific is processing of auditory and visual rhythms? *Psychophysiology*, 53(2), 198–208.
<http://doi.org/10.1111/psyp.12559>
- Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricity and modality on synchronization with a beat. *Experimental Brain Research*, 163(2), 226–238. <http://doi.org/10.1007/s00221-004-2159-8>
- Penner, M. J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *The Journal of the Acoustical Society of America*, 61(2), 552–557.
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential Effects of Auditory and Visual Signals on Clock Speed and Temporal Memory. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1770–1787.
<http://doi.org/10.1037/0096-1523.26.6.1770>
- Petkov, C. I., & Bendor, D. (2016). Neuronal Mechanisms and Transformations Encoding Time-Varying Signals. *Neuron*, 91, 718–721. doi:10.1016/j.neuron.2016.08.006
- Pollack, I. (1968a). Periodicity discrimination for auditory pulse trains. *The Journal of the Acoustical Society of America*, 43, 1113–1119. doi:10.1121/1.1910946
- Pollack, I. (1968b). Detection and Relative Discrimination of Auditory 'Jitter'. *The Journal of the Acoustical Society of America*, 43, 308–315. doi:10.1121/1.1910780
- Ponton, C. W., Moore, J. K., & Eggermont, J. J. (1999). Prolonged deafness limits auditory system developmental plasticity: evidence from an evoked potentials study in children with cochlear implants. *Scandinavian Audiology*. 28(Suppl 51), 13–22.

- Povel, D.-J., & Essens, P. (1985). Perception of Temporal Patterns. *Music Perception: An Interdisciplinary Journal*, 2, 411–440. doi:10.2307/40285311
- Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). “The lower limit of melodic pitch,” *The Journal of the Acoustical Society of America*, 109, 2074–2084. doi:10.2307/4576063?ref=search-gateway:344f817ae4eee114e0c38914d48b9524
- Rammsayer, T. H. (2014). The effects of type of interval, sensory modality, base duration, and psychophysical task on the discrimination of brief time intervals. *Attention, Perception & Psychophysics*, 76(4), 1185–1196. <http://doi.org/10.3758/s13414-014-0655-x>
- Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, 50(6), 565–574.
- Rammsayer, T., & Ulrich, R. (2001). Counting models of temporal discrimination. *Psychonomic Bulletin & Review*, 8(2), 270–277.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89(2), 1078–1093. <http://doi.org/10.1152/jn.00706.2002>
- Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1085–1099. <http://doi.org/10.1037//0096-1523.28.5.1085>
- Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, 68(4), 252–270. <http://doi.org/10.1007/s00426-003-0143-8>
- Ross, J., & Houtsma, A. J. (1994). Discrimination of auditory temporal patterns. *Percept Psychophys*, 56, 19–26.
- Rossi-Pool, R., Salinas, E., Zainos, A., Alvarez, M., Vergara, J., Parga, N., & Romo, R. (2016). Emergence of an abstract categorical code enabling the discrimination of temporally structured tactile stimuli. *Proceedings of the National Academy of Sciences*, 113(49), E7966–E7975. <http://doi.org/10.1073/pnas.1618196113>
- Rothschild, G., Nelken, I., & Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature*, 465, 353–360. doi:10.1038/nn.2484
- Roy, C., Dalla Bella, S., & Lagarde, J. (2017). To bridge or not to bridge the multisensory time gap: bimanual coordination to sound and touch with temporal lags.

REFERENCES

- Experimental Brain Research*, 235(1), 135–151. <http://doi.org/10.1007/s00221-016-4776-4>
- Roye, A., Schröger, E., Jacobsen, T., & Gruber, T. (2010). Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. *Journal of Neuroscience*, 30, 7310–7313. doi:10.1523/JNEUROSCI.1113-10.2010
- Royer, F. L., & Garner, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Percept Psychophys*, 1, 41–47. doi:10.3758/BF03207820
- Saal, H. P., Wang, X., & Bensmaia, S. J. (2016). Importance of spike timing in touch: an analogy with hearing? *Current Opinion in Neurobiology*, 40, 142–149. <http://doi.org/10.1016/j.conb.2016.07.013>
- Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. *Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, 19(22), 10074–10081.
- Salinas, E., Hernandez, A., Zainos, A., & Romo, R. (2000). Periodicity and Firing Rate As Candidate Neural Codes for the Frequency of Vibrotactile Stimuli. *Journal of Neuroscience*, 20(14), 5503–5515.
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature reviews neuroscience*, 10(3), 211–223.
- Sathian, K., & Zangaladze, A. (1997). Tactile learning is task specific but transfers between fingers. *Perception & Psychophysics*, 59(1), 119–128.
- Schnupp, J. W. H. (2006). Plasticity of Temporal Pattern Codes for Vocalization Stimuli in Primary Auditory Cortex. *Journal of Neuroscience*, 26(18), 4785–4795. <http://doi.org/10.1523/JNEUROSCI.4330-05.2006>
- Seiple, W., Greenstein, V., Holopigian, K., & Carr, R. (1988). Changes in the focal electroretinogram with retinal eccentricity. *Documenta Ophthalmologica*, 70(1), 29–36. <http://doi.org/10.1007/BF00154733>
- Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *The Journal of the Acoustical Society of America*, 95(6), 3529–3540.
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123. <http://doi.org/10.1016/j.tins.2010.11.002>
- Shannon, R. V. (1985). Threshold and loudness functions for pulsatile stimulation of

- cochlear implants. *Hearing Research*, 18, 135–143.
- Shannon, R. V. (1989). Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *The Journal of the Acoustical Society of America*, 85(6), 2587–2592. <http://doi.org/10.1121/1.397753>
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303.
- Sherrick, C. E. (1985). A scale for rate of tactual vibration. *The Journal of the Acoustical Society of America*, 78(1), 78–83. <http://doi.org/10.1121/1.392457>
- Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *eLife*, 5, e19113. <http://doi.org/10.7554/eLife.19113>
- Sorkin, R. D. (1990). Perception of temporal patterns defined by tonal sequences. *The Journal of the Acoustical Society of America*, 87(4), 1695–1701.
- Spengler, F., Roberts, T. P., Poeppel, D., Byl, N., Wang, X., Rowley, H. A., & Merzenich, M. M. (1997). Learning transfer and neuronal plasticity in humans trained in tactile discrimination. *Neuroscience Letters*, 232(3), 151–154. [http://doi.org/10.1016/S0304-3940\(97\)00602-2](http://doi.org/10.1016/S0304-3940(97)00602-2)
- Stauffer, C. C., Haldemann, J., Troche, S. J., & Rammsayer, T. H. (2012). Auditory and visual temporal sensitivity: Evidence for a hierarchical structure of modality-specific and modality-independent levels of temporal information processing. *Psychological Research*, 76(1), 20–31. <http://doi.org/10.1007/s00426-011-0333-8>
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52(1), 77–86. <http://doi.org/10.1016/j.ijpsycho.2003.12.005>
- Sutherland, R. J., Wishaw, I. Q., & Kolb, B. (1988). Contributions of Cingulate Cortex to Two Forms of Spatial Learning and Memory. *Journal of Neuroscience*, 8(6), 1863–1872.
- Takeuchi, T., Duzkiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., ... & Morris, R. G. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature*, 537(7620), 357–362.
- Tanner, T. A., Patton, R. M., & Atkinson, R. C. (1965). Intermodality judgments of signal duration. *Psychonomic Science*, 2(1-12), 271–272. <http://doi.org/10.3758/BF03343446>
- Tanner, W. P. (1956). Theory of Recognition. *The Journal of the Acoustical Society of*

REFERENCES

- America*, 28(5), 882–888. <http://doi.org/10.1121/1.1908504>
- Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Front. Psychol*, 5, 1329. doi:10.3389/fpsyg.2014.01329
- Tomassini, A. (2011). Perceived duration of visual and tactile stimuli depends on perceived speed. *Frontiers in Integrative Neuroscience*, 5, 51. <http://doi.org/10.3389/fnint.2011.00051>
- Treisman, M. (1963). Auditory Unmasking. *Journal of the Acoustical Society of America*, 35(8), 1256–1263. <http://doi.org/10.1121/1.1918681>
- Tyler, C. W. (1985). Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions. *JOSA A*, 2(3), 393–398. <http://doi.org/10.1364/JOSAA.2.000393>
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, 8, 1–11. <http://doi.org/10.1038/ncomms14637>
- Van Doren, C. L., Gescheider, G. A., & Verrillo, R. T. (1990). Vibrotactile temporal gap detection as a function of age. *The Journal of the Acoustical Society of America*, 87(5), 2201–2206. <http://doi.org/10.1121/1.399187>
- van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil Dilation Co-Varies with Memory Strength of Individual Traces in a Delayed Response Paired-Associate Task. *PLoS ONE*, 7(12), e51134–8. <http://doi.org/10.1371/journal.pone.0051134>
- van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1815–1830.
- VanRullen, R., Guyonneau, R., & Thorpe, S. J. (2005). Spike times make sense. *Trends in Neurosciences*, 28, 1–4.
- Verrillo, R. T. (1965). Temporal Summation in Vibrotactile Sensitivity. *The Journal of the Acoustical Society of America*, 37(5), 843–846. <http://doi.org/10.1121/1.1909458>
- Verrillo, R. T. (1985). Psychophysics of vibrotactile stimulation. *The Journal of the Acoustical Society of America*, 77(1), 225–232. <http://doi.org/10.1121/1.392263>
- Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., & Lahti-Nuutila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Experimental Brain Research*, 186(4), 525–537. <http://doi.org/10.1007/s00221-007-1254-z>

- Viswanathan, J. (2016). *When noise enlightens: An investigation into the mechanisms of perception and long-term memory for meaningless auditory stimuli* (Doctoral dissertation, S. J. Thorpe, Ed.) Université Paul Sabatier, Toulouse.
- Viswanathan, J., Rémy, F., Bacon-Macé, N., & Thorpe, S. J. (2016). Long Term Memory for Noise: Evidence of Robust Encoding of Very Short Temporal Acoustic Patterns. *Frontiers in Neuroscience, 10*, 610–11. <http://doi.org/10.3389/fnins.2016.00490>
- Vos, P. G., van Assen, M., & Frañek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research Psychologische Forschung, 59*(4), 240–247. <http://doi.org/10.1007/bf00439301>
- Warren, R. M., Bashford, J. A., Cooley, J. M., and Brubaker, B. S. (2001). Detection of acoustic repetition for very long stochastic patterns. *Percept Psychophys, 63*, 175–182.
- Wearden, J. H. (1992). Temporal Generalization in Humans. *Journal of Experimental Psychology: Animal Behavior Processes, 18*(2), 134–144. <http://doi.org/10.1037/0097-7403.18.2.134>
- Wearden, J. H. (2003). Applying the scalar timing model to human time psychology: Progress and challenges. *Time and Mind II: Information Processing Perspectives, 21–39*.
- Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and Generalization of Auditory Temporal–Interval Discrimination in Humans. *Journal of Neuroscience, 17*(10), 3956–3963.
- Yau, J. M., Olenczak, J. B., Dammann, J. F., & Bensmaia, S. J. (2009). Temporal Frequency Channels Are Linked across Audition and Touch. *Current Biology, 19*(7), 561–566. <http://doi.org/10.1016/j.cub.2009.02.013>
- Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research, 174*, 101–106.

Appendix

Neural correlates for time pattern recognition (preliminary results)

This is a short report of an ongoing study investigating neural correlates of auditory memory for random time patterns. The study is in close collaboration with Jennifer Lawlor, Célian Bimbard, PhD students in the team, and Yves Boubenec, in charge of the Neuro platform in the laboratory.

We presented click trains to an awake ferret in a passive listening condition while recording neuronal activity. The click trains were adapted from the study of Kang et al. (2017) as described in Chapter 2. We used click trains with time intervals randomly drawn from a Poisson distribution, either lasting for 3 s (random click trains, C), or 1 s of random click trains immediately repeated 3 times (repeated click trains, RC). One RC with frozen time intervals re-occurred for several trials within a session (reference RC, RefRC). Click rates were either 5 Hz or 12 Hz, with 10 ms of a refractory period, corresponding to a nominal rate of either 5 Hz or 10 Hz. Single-unit recordings were conducted either in the primary auditory cortex (A1) or in the secondary auditory cortex (posterior ectosylvian gyrus, PEG) of the awake but passive ferret. Here we will present preliminary analyses composed of mean spike rates and a decoding analysis applied to local field potentials (LFP) traces.

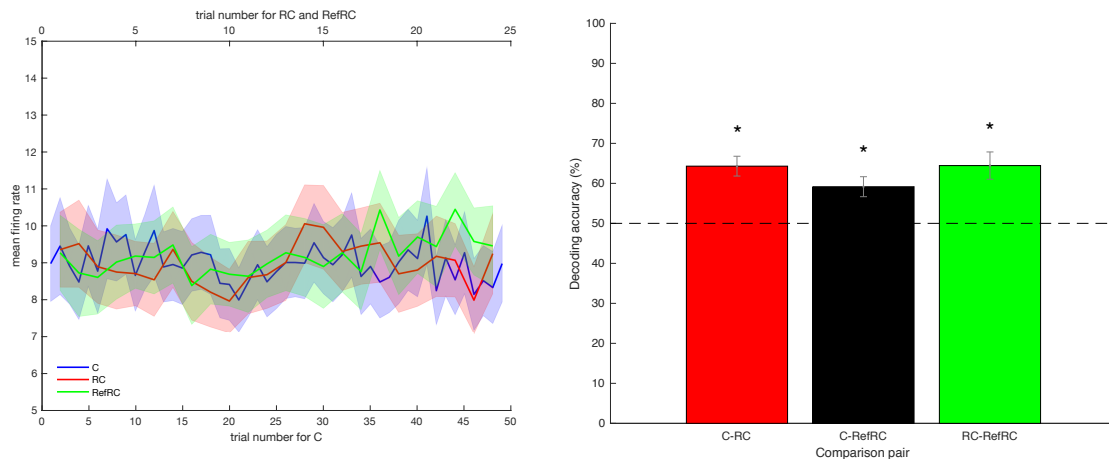


Figure A-1. (Left) Average firing rates across all recorded cells in A1 ($n = 70$, right) for each stimulus type. Bottom x-axis indicates trial order of C (blue) while top x-axis indicates trial order of RC (red) and RefRC (green). Error shades indicate standard error of the mean. (Right) Decoding accuracy at discriminating different stimulus types for each comparison pair using A1 mean spike responses. See text for details. Horizontal dashed lines indicate chance level. Error bars indicate 95% confidence intervals.

In our behavioural study (Kang et al., 2017), we observed an increase in performance with the number of presentations only for RefRC against constant performance for RC. We therefore compared average firing rates for each stimulus type per trial in order to see whether neural responses would show similar divergence across stimulus type or not. We did not find any significant differences between stimulus types across trials ($n = 96$, Figure A-1, left).

Whether neural activities for each condition can be differentiated by more detailed descriptions than simple average rate was further examined by trying to decode the response using linear classification (Fisher, 1936). Population vector arrays of mean firing rates or LFP envelopes for each trial were created to compare different stimulus types: either C vs. RC, RC vs. RefRC, or C vs. RefRC. Half of the trials were used for training the decoder and the rest of the trials were used for testing and computing the decoding accuracy. This cross-validation procedure was repeated 30 times. We used mean firing rates and LFP envelopes in A1 to examine how each stimulus type (C, RC, and RefRC) could be decoded from one another. The obtained accuracy was compared to random permutation results from 100 iterations, to assess statistical significance of the decoding accuracy.

Statistical one-tail z -test confirmed that decoding accuracies for each comparison pair were significantly higher than the accuracy obtained from 100 iterations of the

APPENDIX

random permutation results obtained by shuffling the category labels (C - RC: $z = 2.28$, $p < 0.05$, C - RefRC: $z = 1.74$, $p < 0.05$, RC - RefRC: $z = 3.11$, $p < 0.001$, p -values are uncorrected, Figure A-1, right). Since the first appearance of click trains is always new and fresh regardless of the stimulus type, the discriminability between RefRC to RC or to C should initially be at chance level. The discriminability between stimulus types should then evolve along trials, discriminating RefRC from other sequences. To ascertain whether the distance between stimulus types diverge along the course of re-occurring trials, we projected A1 spike responses on to the decoding classifiers. However, no statistical evidence was found to support this hypothesis.

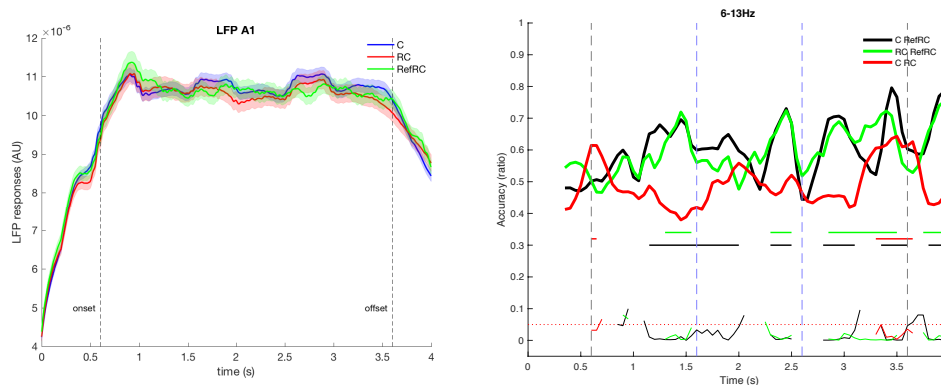


Figure A-2. (Left) *RMS* envelopes of LFP traces for each stimulus type in alpha frequency band (6-13 Hz, blue: C, red: RC, green: RefRC). Vertical dashed lines indicate the onset and the offset of the click trains. Error shades indicate standard error of the mean. (Right) Decoding accuracy of each time point for the stimulus type comparisons in alpha frequency band. Coloured thick solid lines indicate the time course of the decoding accuracy for each comparison (black: C vs. RefRC, green: RC vs. RefRC, red: C vs. RC). Coloured horizontal solid lines indicate the time periods where the significant accuracy ($p < 0.05$, one-tail) was observed for each comparison. Dashed vertical lines indicate the onset and the offset of the click train (grey) and repeated segment onsets within the click train (blue). Dashed horizontal line at the bottom indicates p -value criterion at 0.05. Coloured curved lines at the bottom around the p -value criterion (0.05) indicate time periods where $p < 0.1$.

We then applied the decoding analysis with to LFP traces (Figure A-2, left). We focused on the recordings in A1 since no discriminability was observed in PEG. For LFP traces, the decoding accuracy across different types of stimuli using the LFP *rms* envelopes of each frequency range along the sequence duration was analysed first. From the theta band (4-6 Hz), RefRC against RC and C was discriminable towards the end of

the sequence. In the alpha frequency band (6-13 Hz), better decoding accuracy for RefRC against other stimulus types was observed especially during the third presentation of the segment within the sequence (Figure A-2, right). Furthermore, a high decoding accuracy was found for RefRC against C even from the first presentation of the segment. A significant discriminability between RC and C was also found during the last presentation of the segment, only in the alpha band. We decided to focus on alpha frequency range so that comparisons across all three stimulus types could be further analysed. We wanted to investigate the time-dependent discriminability between stimulus type across trials. We took the time window where all three comparison groups showed significant decoding accuracy to obtain mean weights during that time period (3.3 s to 3.6 s). The LFP envelope values were diverged on the last 300 ms of the third click train in each trial when projected these vectors onto the average decoder in the same epoch, for each comparison pair. This process was repeated trial after trial.

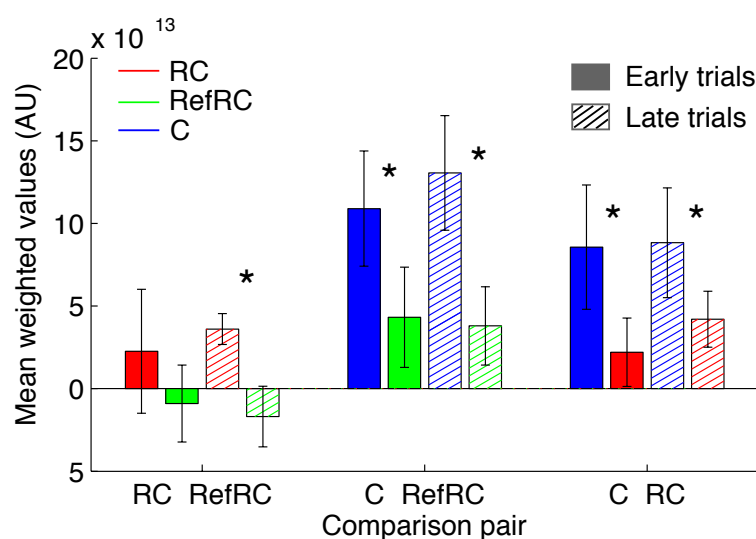


Figure A-3. Mean projected values for early trials (trial 1-10, filled) and last trials (trial 11-24, hatched) per each stimulus type (red: RC, green: RefRC, blue: C) for each comparison pair. For the RC - RefRC comparison, only late trials show statistically significant difference on averaged values between RC and RefRC while averaged values for early trials do not differ. A constant difference was observed for C - RefRC and C - RC comparisons for both early and late trials. Error bars indicate 95% confidence intervals.

For the RC - RefRC comparison, trials were divided into early trials (trial 1-10) and late trials (trial 11-24) for comparisons between RC and RefRC values. Weight

APPENDIX

projection values during early trials did not differ between stimulus types ($t(9) = 1.39, p > 0.05$) but late trials showed significant differences between stimulus types ($t(13) = 5.85, p < 0.001$). Weaker but similar trends were observed for the C - RefRC comparison, although both early and last trials showed significant differences on the projected values between C and RefRC (early trials: $t(9) = 3.70, p < 0.01$, last trials: $t(13) = 4.94, p < 0.001$). Lastly, for the C - RC comparison, the weight projected values were significantly different for early trials and late trials (early trials: $t(9) = 3.06, p < 0.05$, late trials: $t(13) = 2.58, p < 0.05$). The result indicates repeated sounds were always discriminated against non-repeated sounds. In addition, neural responses for later trials allowed to discriminate RC and RefRC, supporting an additional processing encoding a longer-term memory process.

The present finding suggests distinct neuronal activities for the re-occurring time pattern, within the primary cortex in the population response indexed by the LFP. More analyses as well as more recordings with multiple RefRCs within a test session are in progress for further investigation of neural correlates regarding the time pattern recognition.