



HAL
open science

Ambiguïté en perception de hauteur de son : métacognition et effet de contexte

Claire Pelofi

► **To cite this version:**

Claire Pelofi. Ambiguïté en perception de hauteur de son : métacognition et effet de contexte. Psychologie. Université Paris sciences et lettres, 2016. Français. NNT : 2016PSLEE060 . tel-04191805

HAL Id: tel-04191805

<https://theses.hal.science/tel-04191805v1>

Submitted on 30 Aug 2023

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'Ecole Normale Supérieure

Ambiguity in pitch-shift perception: metacognition and context effect

Ecole doctorale n°158

CERVEAU-COMPORTEMENT-COGNITION

Spécialité Sciences Cognitives

**Soutenue par Claire PELOFI
le 14 décembre 2016**

Dirigée par **Daniel PRESSNITZER**

COMPOSITION DU JURY :

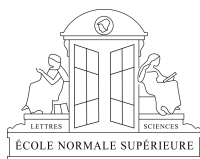
Mme. POULIN-CHARRONAT
Bénédicte
Université de Bourgogne, Rapporteur

M. OXENHAM Andrew
Université du Minnesota, Rapporteur

M. DOKIC Jérôme
Ecole Normale Supérieure, Ecole des
Hautes Etudes en Sciences Sociales,
Membre du jury

M. BEFFA Karol
Ecole Normale Supérieure, Membre du
jury

M. AGON Carlos
IRCAM, Membre du jury



Ecole Normale Supérieure

Ecole doctorale Comportement Cognition Cerveau

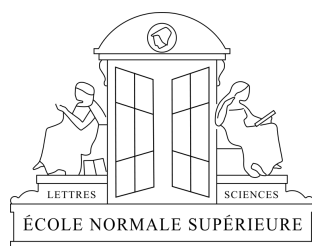
THÈSE

AMBIGUITY IN PITCH-SHIFT PERCEPTION: METACOGNITION AND CONTEXT EFFECT

Claire PELOFI
claire.pelofi@ens.fr

Soutenue publiquement le 14/12/2016 devant un jury composé de :

<i>Rapporteurs</i>	Andrew Oxenham	Univeristé du Minnesota
	Bénédicte Poullin-Charronat	Université de Bourgogne
<i>Examineurs</i>	Jerôme Dokic	École Normale Supérieure
	Karol Beffa	École Normale Supérieure
	Carlos Agon	IRCAM
<i>Directeur de thèse</i>	Daniel Pressnitzer	École Normale Supérieure



Thèse réalisée au Laboratoire des Systèmes Perceptifs
École Normale Supérieure
29 rue d'Ulm
75005 Paris
Tel : +33 1 44 32 26 73
Web : <http://iec-lsp.ens.fr/>

Sous la direction de Daniel Pressnitzer
daniel.pressnitzer@ens.fr

Financement Allocation de recherche LabEx IEC, ENS

Abstract

Perceptual systems are endorsed with the challenging task to organize a partial sensory input into a useful and stable representation of the world. In the auditory modality this task is named "auditory scene analysis" and consists in grouping or streaming the acoustic input into different voices. However, perceptual systems have to confront the fact that perception is an "ill-posed problem": the poverty of signal does not allow to determine the state of the acoustic scene. However, this inherent ambiguity is being processed unbeknownst to listeners. Perceptual systems rely on different strategies to settle on the most adjusted representation of the world, taking into account past information and prior knowledge.

Ambiguous stimuli are experimental objects designed to highlight the processes involved when disambiguating a signal. The aim of this Ph.D. has been to investigate the way listeners process an ambiguous auditory stimulus. The stimulus consists in pairs of tones composed of octave-related frequency components, also known as Shepard tones. When two Shepard tones are separated by a six semi-tones interval, the pitch-shift is ambiguous: listeners report an "upward" and a "downward" pitch-shift with equal likelihood. However, when the ambiguous interval is preceded by a short context melody, listeners are strongly biased towards one or the other pitch-shift: the reported direction encompasses the frequency region where the context components are placed. My aim has been first to characterize the time course of this context effect, for it conveys informations about the nature of underlying neural processes. The results revealed remarkable temporal

dynamics, as the context effect was established by extremely short tones (of the order of 20ms) and yet could still bias perception across long periods of silence (of the order of 30s to minutes).

Secondly, the research tackled a novel question on perceptual ambiguity: are listeners aware of the existence of the distinct perceptual alternatives? Oddly, this question has so far raised relatively small interest, perhaps because the "exclusivity of alternative percepts" has been traditionally considered one of the hallmarks of perceptual ambiguity. In the case of the ambiguous interval, an effect of musical expertise has been discovered. Collecting confidence ratings and response times, it was shown that the pattern of response for these two measures diverged between musicians and non-musicians. A maximally ambiguous interval (corresponding to a 0.5 ratio between "up" and "down" responses) was associated to lower confidence ratings and slower response times for musicians while the opposite pattern was observed for non-musicians. An online experiment replicated these findings on a large cohort of participants and showed a positive correlation between the perceived ambiguity and the number of years of musical training. Finally, two experiments were conducted on the effect of context information on perceived ambiguity, using confidence ratings and response times. Results demonstrate that the context effect strength is paralleled by a reduced perceived ambiguity.

This research sheds light on the perception of an ambiguous auditory interval that is reported as going "up" or "down" with equal likelihood. The temporal dynamics of a context effect were investigated and suggest that similar mechanisms may be involved in the processing of daily auditory scene, which typically require integrating information over wide temporal scales, hence emphasizing the ecological relevance of this work. Finally, a novel approach to perceptual ambiguity was proposed, by combining response times and confidence ratings, in order to tackle the largely unexplored question of the subjective experience of perceptual ambiguity.

Keywords: ambiguity, pitch, context effects, musicianship

Résumé

La tâche des systèmes sensoriels est de d'organiser l'information lacunaire collectée par les organes sensoriels afin d'aboutir à une représentation stable et cohérente du monde. Dans la modalité auditive, « l'analyse de scène auditive » consiste à regrouper ou séparer correctement le signal en différentes voix. Mais les systèmes perceptifs doivent prendre en compte le fait que la perception est un « problème mal-posé » : le caractère lacunaire du signal ne permet pas de reconstituer entièrement la scène. Or, pour des raisons écologiques évidentes, cette ambiguïté doit être traitée très rapidement et à l'insu de l'observateur. Les systèmes perceptifs adoptent alors plusieurs stratégies afin de construire une représentation du monde pertinente et stable en tenant compte de l'information passée et des connaissances *a priori* sur le monde.

Les stimuli ambigus sont des objets expérimentaux conçus pour mettre en évidence ces stratégies perceptives. Ce travail de doctorat s'est attaché à explorer les processus liés au traitement d'un objet ambigu de la modalité auditive. Il s'agit d'un intervalle formé par la succession de deux sons de Shepard, des composés complexes de séries d'octaves. Quand ils sont séparés d'un intervalle de six demi-tons, la perception est ambiguë : les auditeurs rapportent que l'intervalle monte et descend avec une probabilité égale. En revanche, quand l'intervalle est précédé d'une courte mélodie, la perception est fortement influencée : la direction rapportée est celle qui traverse les composantes spectrales de la mélodie qui précède.

Le but de cette travail de thèse a été tout d'abord de caractériser les proprié-

tés temporelles de cet effet de contexte, car elles sont susceptibles d'apporter des indications sur la nature des processus neuronaux sous-jacents. Les résultats révèlent des dynamiques remarquables : l'effet du contexte est établi avec un son contexte de 20 ms seulement et peut être maintenu au delà de 30 s de silence. Par conséquent, on peut supposer que des processus similaires sont en jeu lors du traitement de scènes auditives nécessitant l'intégration de l'information sur de larges échelles de temps, comme lors de la perception de la parole ou de la musique. L'étude des effets de contexte d'un intervalle ambigu constitue donc un outil nouveau et performant pour appréhender les processus sous-jacents lors de la perception de scènes auditives naturelles.

Dans un second temps, mes recherches ont porté sur l'expérience subjective de l'ambiguïté : est-on conscient de l'existence d'alternatives différentes lors du traitement de stimuli ambigus ? Etrangement, cette question semble n'avoir suscité que peu de recherches jusqu'à présent, la raison étant peut-être que l'exclusivité mutuelle des alternatives est traditionnellement considérée comme l'une des caractéristiques de l'ambiguïté perceptive.

Dans le cas de l'intervalle ambigu, un effet de l'expertise musicale a été mis en évidence. En recueillant les jugements de confiance et les temps de réponse associés au jugement de direction, nous avons montré que ces mesures de l'ambiguïté ressentie divergeaient drastiquement entre les musiciens et les non-musiciens. Des jugements de confiance plus bas et des temps de réponse plus longs étaient associés à une ambiguïté de réponse maximale (équiprobabilité des réponses « monte » et « descend ») chez les musiciens alors que l'effet inverse était observé chez les non-musiciens. Une expérience menée en ligne a montré que cet effet de l'ambiguïté était robuste lorsqu'une large cohorte était testée et positivement corrélé aux nombres d'années de pratique musicale. Enfin, deux expériences ont été menées sur l'effet du contexte et l'ambiguïté perçue, en utilisant les mesures de confiance et de temps de réaction. Les résultats ont permis d'établir que pour les deux groupes, un effet du contexte plus fort était associé à une ambiguïté moins ressentie.

Ces travaux ont donc contribué à mettre en lumière les caractéristiques de

la perception d'un objet auditif ambigu, un intervalle pouvant être rapporté comme montant ou comme descendant. Les dynamiques d'un effet de contexte ont été mises en évidence et suggèrent que des mécanismes similaires sont en jeu lors du traitement de la plupart des scènes auditives complexes, ce qui souligne la pertinence écologique de ces travaux. Enfin, une approche nouvelle des stimuli ambigus, alliant mesures de confiance et temps de réponse, a été développée afin d'explorer un aspect jusqu'alors peu documenté de la perception ambiguë, l'expérience subjective de l'ambiguïté.

Mots clés : ambiguïté, hauteur, effet de contexte, expertise musicale

Remerciements

Je souhaiterais tout d'abord remercier chaleureusement Daniel Pressnitzer pour la confiance qu'il m'a témoignée dès mon arrivée dans son équipe. Je sortais alors de mes études de philosophie esthétique et je commençais mes premières semaines de Cogmaster, bien décidée à ne pas laisser mon parcours hybride mais tout de même plutôt littéraire entraver mes vellités de recherche en sciences cognitives. Daniel a répondu avec bienveillance et enthousiasme à mes propositions, lorsque je me suis présentée dans son bureau avec l'envie de travailler "sur la perception des sons". J'ai commencé à travailler sous sa direction en stage et ces travaux ont ensuite menés à la thèse que je présente ici. Tout au long de cette collaboration, Daniel s'est montré attentif au développement de mon indépendance scientifique, et m'a laissé toute latitude pour développer le projet de recherche qui correspondait à mes aspirations et à mon profil. Ce projet nous a conduit sur le chemin de Paul Egré, que je souhaite remercier également avec la plus grande chaleur. La collaboration avec Paul m'a permis de développer une approche conceptuelle des résultats que nous collectons en laboratoire. Le travail que nous avons mené ensemble m'a permis mettre en pratique cette idée que la philosophie et le travail conceptuel peut et doit s'insinuer partout et chaque jour dans le travail d'un chercheur en sciences. Merci également à Vincent de Gardelle pour sa généreuse collaboration et l'apport de ses connaissances. Les heures que nous avons passé à travailler ensemble sur l'analyse de nos données m'ont également montré qu'en code comme en bien des choses, ma marge de progression était encore bien vaste.

Je souhaite remercier Claire Chambers pour son infinie patience et ses encouragements quotidiens lors de mes débuts avec Matlab, R, Python, Tex et autres langues étrangères qu'elle m'a aidé à apprendre. Merci à Shihab pour l'infinie générosité dont il fait a preuve à mon égard, pour les joyeuses conversations que nous eut sur des sujets plus ou moins scientifiques... Pour son inépuisable énergie, pour les programmes de France Musique chaque matin, le fromage et les idées de balade dans Paris. Je souhaite remercier aussi Christian Lorenzi pour la bienveillance, les noisettes et polyphonies corses et la poésie grecque. Merci aussi à mes merveilleux compagnons de laboratoire, Agnès, Dan, Trevor, Romain, Victor, Marion, Bernard, Yves, Thibault, Jennifer, les jolies petites créatures aux noms de fromage et tous ceux que j'e n'ai pas la place de nommer ici. J'adresse une pensée à mon camarade joyeux et plein de fantaisie, Jonathan, né comme moi un jour de paix, et parti bien trop tôt.

Merci à tous mes participants, sans les oreilles de qui, ne l'oublions pas, rien de tout cela n'aurait pas été possible.

Je remercie également ma famille. Je salue tout particulièrement le flegme de ma mère et son obstination têtue à me soutenir dans les moments les plus critiques comme les plus agréables. Je pense également à ma grande tante qui n'a jamais cherché à comprendre le pourquoi du comment de mes recherches, mais qui m'a tout de même soutenu par sa présence fidèle et discrète. Merci à mes amis qui me soutiennent et me supportent depuis tant d'années. Merci à Florian pour ses remarques percutantes et éclairées sur «les ondes Shepard», merci à mes chères Hannah, Elsa, Sonia, Mélisse et Agathe pour être à mes cotés depuis si longtemps et m'avoir soutenu pendant cette thèse.

Merci enfin à Andromaque qui n'a jamais failli à son rôle de chauffe-pied ronronnant, durant les longs mois de rédaction de ce manuscrit. Et merci enfin de tout coeur à Beltran qui partage ma vie depuis le tout début de cette aventure, et qui a été présent, serein et confiant, à chaque étape, chaque doute, chaque victoire et chaque espoir.

Contents

Abstract	i
Résumé	iii
Remerciements	vii
Contents	ix
List of Figures	xv
List of Tables	xxxi
Introduction	1
PART I : Literature review	13
1 Ambiguity in perception	15
1.1 An hint of historical perspective	15

CONTENTS

1.2	The main paradigms	16
1.2.1	Binocular rivalry	16
1.2.2	Bistable visual figures	17
1.2.3	Word repetition	18
1.2.4	Ambiguous auditory streaming	19
1.2.5	The Shepard paradigm	22
1.3	General properties of perceptual ambiguity	26
1.3.1	Similar behavioral observations	26
1.3.2	Bottom-up and top-down integration	27
1.3.3	The special case of exclusivity	35
1.4	Ambiguity across individuals: inter-subject variability	44
1.4.1	Behavioral evidences for individual differences	45
1.4.2	Genetic explanations	48
1.4.3	Neurophysiological explanations	50
2	Context effects and perceptual ambiguity	53
2.1	Simultaneous context	53
2.2	Contrastive vs. assimilative	54
2.2.1	Distinct functional benefits	55
2.2.2	Auditory context effects	56

2.3	Temporal dynamics of context effects	63
2.3.1	Long context durations	63
2.3.2	Short context durations	64
2.3.3	Gap duration	67
2.3.4	Systematic examination of temporal dynamics	69
3	To be or not to be a musician	73
3.1	Makes (almost) no difference	76
3.1.1	Musical tension	77
3.1.2	Musical expectancies	79
3.2	Makes a (big) difference	84
3.2.1	Anatomical differences	85
3.2.2	Behavioral differences	87
	PART II : Experimental contribution	95
4	The temporal dynamics of an auditory context effect	97
4.1	Experiment 1.1	101
4.1.1	Method	101
4.1.2	Results	106
4.1.3	Summary	108

CONTENTS

4.2	Experiment 1.2	109
4.2.1	Method	109
4.2.2	Results	112
4.2.3	Summary	114
4.3	Experiment 1.3	115
4.3.1	Method	115
4.3.2	Results	119
4.3.3	Summary	125
4.4	Discussion	126
5	Ambiguity is perceived differently by musicians and non-musicians	131
5.1	Experiment 2.1	135
5.1.1	Method	136
5.1.2	Results	140
5.1.3	Summary	150
5.2	Exp. 2.2: online experiment	150
5.2.1	Method	151
5.2.2	Results	153
5.2.3	Summary	160
5.3	Experiment 2.3	162
5.3.1	Method	163

5.3.2	Results	165
5.3.3	Interim discussion	169
5.4	Discussion	169
6	Context effect and perceived ambiguity: when musicians and non-musicians accord	175
6.1	Experiment 3.1	178
6.1.1	Method	178
6.1.2	Results	181
6.1.3	Summary	186
6.2	Experiment 3.2	187
6.2.1	Method	188
6.2.2	Results	190
6.2.3	Summary	195
6.3	Discussion	196
	General discussion	199
	Bibliography	207
	Appendix	251
A	Shepard tones and context effects	253

CONTENTS

B Effect of very short context tone	255
B.0.1 Method	255
B.0.2 Results	257
B.0.3 Discussion	258
C Can distraction tasks interrupt context effect in Shepard tone paradigm?	259
D Model of pre-perceptual grouping	261

List of Figures

0.0.1	Example of figure of a smiling face derived from <i>Gestalt</i> principles of perceptive organization	2
0.0.2	Perception is an ill-posed problem: the equation has too many unknowns and too few knowns to be resolved.	3
0.0.3	Context effect in visual perception. The central dot in the two figures seems to have a different size due to the surrounding dots although they are identical. From [378].	5
0.0.4	Prior knowledge effect in visual perception. The left dot seems convex while the second seems concave.	5
0.0.5	Jastrow's ambiguous figure of a duck-rabbit. Here can be seen either a rabbit figure or a duck figure. This example was then taken by Wittgenstein to lead his thoughts on general perception [406].	7
0.0.6	Escher's stairs	9
0.0.7	Ligeti's <i>Evil stairs</i>	10
1.2.1	Procedure of binocular rivalry. Two different images are presented to both eyes. From [360].	17

LIST OF FIGURES

1.2.2	Illustrations of ambiguous stimuli. Left panel: Necker cube, from [360]. Middle panel: Rubin vase/face figure, from [371]. Right panel: schematic representation of an ambiguous apparent motion known as the « spinning wheel illusion », from [360].	18
1.2.3	Word repetition paradigm. Figure by Agnès Leger from [334].	18
1.2.4	Results of a verbal transformation paradigm. List of the possible interpretations for the repeated word « Lame-duck » and « Tress ». From [393].	19
1.2.5	Illustration of the streaming paradigm. From [360].	20
1.2.6	Ambiguity region in the streaming paradigm. From [257].	21
1.2.7	Helical representation of pith along the dimension of height and chroma, or class. The highest in the helix a tone is, the highest the corresponding perceived pitch. The figure also captures the fact that two tones situated on the same vertical axis on the helix will have close pitch chroma. From [343].	23
1.2.8	One Shepard tone. Each vertical line represent a frequency component whose amplitude is determined by a Gaussian envelope.	24
1.2.9	Proximity principle in perceived pitch-shift direction of Shepard tones intervals.	25
1.3.1	Histograms of dominance durations for the auditory (top) and visual (bottom) modalities. There is no significant difference between the two distributions. From [295].	27

1.3.2	Intra-individual regression analysis. Each circle corresponds to an individual point of signal intensity as a function of the number of perceptual transitions in the left anterior cingulate cortex (ACC) (closed circles) and left inferior frontal cortex (IFC) (open circles). From [202].	30
1.3.3	The central item can be perceived as the number "13" or the letter "B" depending on which informations is surrounding it. From [51].	32
1.3.4	Integrative modal of lateral inhibition and excitatory feedback projections. From [379]	33
1.3.5	Experimental design and results from Intaite et al. [165].	34
1.3.6	Results of Miller et al. Example of mixture percept report in binocular rivalry procedure. From [251].	36
1.3.7	Results of Klink et al. Left panel displays the possible alternative percepts that observers could report. The right panel corresponds to the proportion of exclusive percepts over time averaged across participants and computed for 100s stimulation. From [188].	37
1.3.8	Example of soritical series of color shades gradually changing from a clear green to a clear blue. From [105].	38
1.3.9	Results from Grinband et al.. Linear regression between uncertainty and response time averaged across participants (n=10), R=0.342, (error=SD). From [140].	39
1.3.10	Stimulus used in the first experiment in Takei & Nishida. From [371].	40
1.3.11	Normalized reaction times (RT) plotted against response ambiguity. Each point corresponds to a participant RT and response ambiguity for each five disparity conditions. From [371].	41

1.3.12 Schematic illustration of the two conditions (unambiguous and ambiguous) in experimental method. From [168].	42
1.3.13 Mean reaction times results. Each plot corresponds to individual data. Black triangles: Unambiguous stimuli. Gray squares: Ambiguous stimuli. Arrows marked L and R underneath the horizontal axis (top) show the direction of motion in the left and right eyes, respectively, in the ambiguous condition. Solid lines show the model fit. From [168].	43
1.4.1 Intra-individual regression analysis. Each circle corresponds to an individual point of signal intensity as a function of the number of perceptual transitions in the left anterior cingulate cortex (ACC) (closed circles) and left inferior frontal cortex (IFC) (open circles). From [202].	46
1.4.2 Results from experiment 1 in Denham et al.. Differences between transition matrices from the same participant at different sessions (black dots; median—black diamonds, line), differences between the transition matrices of each participant and those of all other participants (red dots; median—red diamonds, line). From [86].	47
1.4.3 Results of Miller et al.. Correlation coefficients are compared between MZ (black bars) and DZ (white bars) twin pairs. The pairs of bars correspond to perceptual switch rate, predominance time, number of mixed percept reports and time spent on mixed percept respectively (error bars indicate 95% CI). From [251].	49
1.4.4 Results of Kanai et al.. Individual standardized grey matter density in the central voxel of the region of interest are plotted against percept durations. Each dot corresponds to a participant (n=52). From [170].	50

2.2.1	Test sequence and context sequence in a study on contrastive effects in vowel identification. From [216].	57
2.2.2	Left panel: schematic representation of experimental procedure: acoustic history with distinct frequency range is followed by a speech stimulus after 50ms of silence. Right panel: mean percentage of "ga" response in each test stimulus condition (consonant-vowel condition), for each context condition (different curves). From [158].	58
2.2.3	Individual results of two listeners. The proportion of "left" response is plotted against the ILD condition of the test tone for the different context conditions. From [75].	59
2.2.4	Averaged proportion of "falling" reported direction plotted as a function of the onset frequency of the test tone, for each context condition (different curves) and each frequency range of context (different panels). From [392].	60
2.2.5	Left panel: schematic representation of increasing and decreasing order of intervals between standard (black) and comparison (grey) Shepard tones (y-axis in frequency log). Right panel: averaged proportion of standard tone perceived as higher than comparison tone as a function of the interval size for the increasing and decreasing order conditions. From [63].	62
2.3.1	Proportion of visual motion priming as a function of the duration of context presentation. Each line corresponds to a different spatial-frequency condition of grating and each plot corresponds to a different contrast condition. From [286].	65

LIST OF FIGURES

2.3.2	Inter-trial procedure in a spectral motion study. Listeners are asked on each trial to report the direction of the glide. A current trial (t) is both the context of the subsequent trial (t+1) and the test of the preceding trial (t-1). Colors code the context condition for each trial: red, downward context, black, flat context, blue, upward context. From [6].	66
2.3.3	Proportion of "ba" responses following a context sequence of various durations and as a function of the frequency of the target stimulus. From [392].	67
2.3.4	Perceptual alternations in continuous presentation conditions and perceptual stabilization in intermittent presentation conditions. From [44].	67
2.3.5	Proportion of time trials heard as streaming for the context period (left portion of panel) and test period (right portion of panel). Trials are sorted according to whether they ended on a two-stream percept (dashed line) or a one-stream percept (solid line). From [352].	69
2.3.6	Left panel displays the stimuli and procedure in Kanai et al. Right panel displays the results. The proportion of motion congruency between context and test stimuli are displayed as a function of different ISI (on the x-axis) and for different context durations (different lines). From [172].	70
2.3.7	Amount of contrastive (suppression) or assimilative (facilitation) effect as a function of context presentation duration for different contrast levels. Brief exposure enhances its subsequent dominance (facilitation), but exposure durations over about .5 s have the opposite effect (suppression). From [45].	71
3.1.1	From [33].	78

3.1.2	Example of two different sequences preceding two chords, one facilitate the processing of the target chords while the other does not. From [28].	80
3.1.3	Percentage of correct response (left panel) and response time (right panel) averaged across sequences for the two target chord conditions (dissonant in dashed line and consonant in solid line) and for the two groups, musicians and non-musicians. From [28].	80
3.1.4	Sensory vs. cognitive account of harmonic priming. From [32].	81
3.1.5	Top panel displays examples for syntactically correct chord progression in sequence (i) and incorrect one in sequence (ii). The bottom panel displays the ERPs elicited by the final chords of these two sequence types (recorded from a right-frontal electrode site from 12 subjects. From [192].	83
3.1.6	Difference in ERPs (expected chords subtracted from unexpected chords) elicited in musicians (solid line) and non-musicians (dotted line). From [193].	83
3.2.1	Brain surface projection of the right and left <i>planum temporale</i> in a musician (top panel) and a non-musician (bottom panel). From [327].	86
3.2.2	Correlations between participant's total amount of practice and white matter structure. Individual points display the mean FA (index of white matter organization) value across voxels for each participant. Vertical lines represent s.d. within that participant. Dashed lines are regression lines. From [18].	87

LIST OF FIGURES

3.2.3	Points indicate averaged individual frequency threshold as a function of the number of years of musical practice for each reference frequency. Curves indicate best fitting exponential curves. From [182].	88
3.2.4	An idealized categorical speech perception function. The stimulus covers a continuous range and yet, perception dramatically shift from one category to the other at a particular point in the continuum. From [278].	89
3.2.5	The hierarchical structure of auditory processing, from cochlea to primary auditory cortex. Red connections indicate bottom-up and top-down neural connections responsible for enhanced low-level encoding of complex signals for musically trained listeners. From [281, 181].	91
3.2.6	The top plots display FFR waveforms for musician and non-musician listeners. The bottom plots displays the fundamental frequency of the speech signal as the (black line) and the FFR's primary periodicity (orange line) for musicians and non-musicians. The tracking of fundamental frequency is more accurate for the musician listener. From [407]. . .	92
4.1.1	Shepard tones were generated by adding octave-related pure tones, weighted by a bell-shaped amplitude envelope. . . .	102
4.1.2	In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is, that corresponds here to an "upward" pitch-shift. . . .	103

4.1.3 The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean and lines indicate individual results. 107

4.2.1 In a trial, a context sequence composed of five tones preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. The duration of the silent gap between the context sequence was varied. 111

4.2.2 The proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean and lines indicate individual results. 113

4.3.1 In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. The duration of the tone and the silent gap between the context sequence were varied. 116

4.3.2 Each panel displays the proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners for one context duration. Shaded areas indicate ± 1 standard error about the mean. 119

4.3.3 The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Bars indicate ± 1 standard error about the mean. Post-hoc multiple comparisons of means for pairs of context duration factor. Only pairs with a significant difference are reported. 122

LIST OF FIGURES

4.3.4 Top panel: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners. Bars indicate ± 1 standard error about the mean. Bottom panel: Post-hoc multiple comparisons of means for pairs of gap duration factor. Only pairs with a significant difference are reported. 123

4.3.5 The proportion of “biased” responses, $P(\text{bias})$, is displayed for each C-T1 onset delays and averaged across listeners in real scale. Bars indicate ± 1 standard error about the mean. 124

4.3.6 The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Bars indicate ± 1 standard error about the mean. 125

5.1.1 Shepard tones were generated by adding octave-related pure tones, weighted by a bell-shaped amplitude envelope. 137

5.1.2 Pairs of Shepard tones, T1 and T2 for thee interval conditions, 3 st, 6 st and 9 st. 138

5.1.3 The proportion of “up” responses, $P(\text{up})$, is displayed for each interval and averaged within groups (non-musicians: red o, musicians: blue x). For all panels, shaded areas indicate ± 1 standard error about the mean. The interval of 6 st corresponds to the ambiguous case. 141

5.1.4 Individual results for non-musicians: in red, the proportion of “up” responses, $P(\text{up})$ is displayed for each interval and in black, the fitted psychometric curve. 142

5.1.5 Individual results for musicians: in blue, the proportion of “up” responses, $P(\text{up})$ is displayed for each interval and in black, the fitted psychometric curve. 143

5.1.6	Response times for each interval, averaged within groups (non-musicians: red o, musicians: blue x). The natural logarithms of RTs were used to compute mean and standard error; y-labels have been converted to milliseconds for display purposes.	144
5.1.7	Ambiguity effect for log-RTs is obtained by averaging the log-RTs for non-ambiguous intervals (all intervals in the 1 st – 11 st range except 6 st) and comparing this value with the log-RT at 6 st within each group (non-musicians: red o, musicians: blue x). Bars indicate +-1 standard error about the mean and individual symbols indicate individual results.	145
5.1.8	Confidence ratings on a scale to 1 (very unsure) to 7 (very sure) for each interval, averaged within groups (non-musicians: red o, musicians: blue x).	146
5.1.9	Ambiguity effect for confidence is obtained by averaging the confidence for non-ambiguous intervals (all intervals in the 1 st – 11 st range except 6 st) and comparing this value with the confidence at 6 st within each group (non-musicians: red o, musicians: blue x). Bars indicate +-1 standard error about the mean and individual symbols indicate individual results.	147
5.1.10	Meta-d'/d' for both groups (non-musicians: red o, musicians: blue x). Bars indicate +-1 standard error about the mean and individual symbols indicate individual results. . .	148
5.1.11	Correlation between confidence and log-RT. Each point represent an interval condition for an individual listener, solid lines are fitted linear regressions over all intervals for each individual listener (non-musicians: red o, musicians: blue x).	149
5.2.1	Histogram of the number of years of training in instrument from participants	154

LIST OF FIGURES

5.2.2 The thick lines represents the averaged proportion of "up" responses across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function each interval. The shaded area corresponds to the standard error. 155

5.2.3 The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of each interval. The shaded area corresponds to the standard error. 156

5.2.4 Histograms of the ambiguity effect, in confidence units, for non-musicians (red) and musicians (blue). The ambiguity effect was defined as the average confidence for all intervals excluding 0 st and 6 st, minus confidence for 6 st, the ambiguous case. Negative values correspond to listeners being less confident for the ambiguous case. 157

5.2.5 Correlation between the ambiguity effect and years of formal musical training. Blue stars indicate participants aged 35 yrs or less. Solid lines indicate the linear regression (blue line) and 95% confidence interval (red lines) fitted to the data for these younger participants. 159

5.2.6 Each bar correspond to a the XX averaged across participants in both groups, in red, non-musicians and in blue, musicians. Error bars correspond to standard error. 161

5.2.7 Correlation between the degree of musical expertise, measured as the number of years of instrument training, and the XX in the mistuning task. 161

5.3.1 Schematic representation of stimuli in Exp. 2.1. Each frequency component in T1 and T2 is "jittered". 164

5.3.2 Averaged "up" response averaged across participants as a function of T1-T2 interval. The different plots correspond to the baseline and each jitter condition (error = SE). . . . 166

5.3.3 Averaged confidence ratings averaged across participants as a function of T1-T2 interval for the baseline and each jitter condition (error = SE). 167

5.3.4 For each participant, we normalized the ambiguity effect observed in each condition by subtracting the ambiguity effect observed for the baseline. The panel displays individual means and standard errors about the means. Negative values signal a dip in confidence for the ambiguous case, when jitter was applied. 168

6.1.1 In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. Here it would result in a upward shift response. The duration of the context tone was varied. 179

6.1.2 The thick lines represents the context bias averaged across listeners and plotted as a function of context duration conditions. In red are non-musicians and in blue, musicians. The shaded area corresponds to the standard error. . . . 181

6.1.3 The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of the context duration condition (shaded areas correspond to standard error), from Z-score data (left panel) and raw data (right panel). 183

LIST OF FIGURES

6.1.4 Individual linear regression between confidence and P(bias). In blue, individual musician data and in red, individual non-musician data. 184

6.1.5 Averaged logarithmic response times across participants in the two groups (red: non-musicians, blue: musicians) are plotted as a function of the context duration condition (error = SE) from Z-score data (left panel) and raw data (right panel). 185

6.1.6 Top panel: Correlation between log-RTs and context bias strength. In blue, individual musician data and in red, individual non-musician data. 186

6.2.1 Results from a previous experiment (see Appendix A, section "Context Effect in Pitch Perception"). Effect of frequency of the context tone. The mean P(T1 lower) is shown as a function of the interval between C and T 1 (error bars = SE).188

6.2.2 Schematic representation of Exp 3.2 design. The duration of the context tone was fixed and the frequency was varied between 0 st and 12 st relative to T1. 189

6.2.3 The thick lines represents the context bias averaged across listeners and plotted as a function of context frequency conditions. In red are non-musicians and in blue, musicians. The shaded area corresponds to the standard error. . . . 191

6.2.4 The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of the context frequency condition (shaded areas correspond to standard error), from Z-score data (left panel) and raw data (right panel).192

6.2.5 Individual linear regression between confidence and P(bias). In blue, individual musician data and in red, individual non-musician data. 193

6.2.6	Averaged log-RTs across participants in the two groups (red: non-musicians, blue: musicians) are plotted as a function of the context duration condition (error = SE).	194
6.2.7	Individual linear regression between log-RTs and P(bias). In blue, individual musician data and in red, individual non-musician data.	195
B.0.1	The proportion of “biased” responses, P(bias), is displayed for each context duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean.	258

List of Tables

4.1	Post-hoc analysis results are reported. T-tests with a Bonferroni correction were conducted and the proportion of "biased" response $P(\text{bias})$ was compared to 0.5 in each context condition. Reported p -values are multiplied by the number of observations $N = 8$	108
4.2	Post-hoc analysis results are reported. t -tests with a Bonferroni correction were conducted and the proportion of "biased" response $P(\text{bias})$ was compared to 0.5 in each gap condition. Reported p -values are multiplied by the number of observations $N = 8$	114
4.3	Post-hoc multiple comparisons of means for pairs of gap and context duration factor. Only pairs with a significant difference are reported.	121
5.1	Mean and standard error of ambiguity effect on R_t -logs for the two groups.	146
5.2	Mean and standard error of ambiguity effect on Confidence ratings for the two groups.	147

LIST OF TABLES

6.1	The averaged context bias across all participants (both groups mixed) for each condition is compared with 0.5. The reported p -values are multiplied by the number of observations $N = 8$	182
-----	--	-----

"Da wird für eines Augenblickes Zeichnung,
ein Grund von Gegenteil bereitet, mühsam,
daß wir sie sähen; denn man ist sehr deutlich
mit uns. Wir kennen den Kontur
des Fühlens nicht: nur, was ihn formt von außen"

"Then briefly a design that's based on contrast,
comes into focus, carefully prepared
for us to see. (They take some pains with us.)
We do not know the contour of our feeling:
only the thing that moulds it from without"

"Pour nous permettre de voir le dessin d'un instant,
on lui prépare laborieusement
un fond de contraste ; car nous devons être très clairs envers nous-mêmes.
Le contour du sensible, nous ne le connaissons pas : seulement
ce qui du dehors le façonne"

Rainer Maria Rilke

Introduction

What does perception?

Perceptual objects

As soon as we enter the world, our senses collect a tremendous amount of information from which they have to derive a useful representation of the environment in order to adjust behaviors. Grouping and segregating elements into different perceptual objects is what constitute the core of this processing. In the auditory modality, this is named "auditory scene analysis"[\[47\]](#) and consists in grouping the parts of the spectrogram that form one "auditory object" or "stream" and segregating the parts of the spectrogram that are emitted by distinct sources. Being critical to survival, infants from a very early age are able to segregate information coming from the environment in order to form auditory objects [\[82\]](#) and this ability is sharpened progressively throughout development [\[241, 367\]](#).

In the 1930s, Gestalt psychologists gathered the principles underlying the formation of « behavioral units ». To illustrate the mechanisms of such principles at play, consider Figure 0.0.1. Having one glance at it, nobody could refrain seeing a smiling figure. Yet, the figure also depict eleven dots in a certain spatial arrangement without any connection between them.

This principle of grouping together elements according to proximity cues constitutes one of the fundamental aspects of perception. Critically, it is



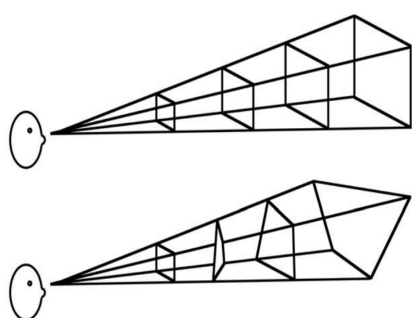
Figure 0.0.1: Example of figure of a smiling face derived from *Gestalt* principles of perceptive organization

the organization *as it appears to us* that matters. When gazing at a starry sky, the factual relations of distance between the objects are of the least importance when grouping them into visual objects: only the relations as they appear to eyes will lead the observer to form constellations. According to Koffka, real units and their relations are neither the necessary nor the sufficient condition to form behavioral units [199]. By the time they were theorized, *Gestalt* principles were mostly applied to visual perception but as Bregman claimed: "in general, all the gestalt principles of grouping can be interpreted as rules for scene analysis" [47].

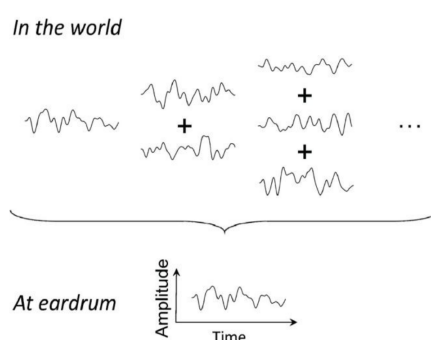
Ill-posed problem

Perception is about forming sensible and stable perceptual objects (visual objects and auditory streams) that can lead our behaviors into a fitted action. Yet, the sensory input coming through the sensors is inherently partial in two ways. First the information about a particular object is rarely entirely available in the world. Consider a cat chasing a mouse in the grass, its tracking of the visual object "mouse" might be oftentimes obstructed by grass stalks. Equally, in a restaurant, when having lunch in a crowded room, a listener might miss parts of his interlocutor's speech because of the concurrence of an other nearby speaker's voice [398]. Therefore, perceptual systems have to overcome the simultaneous masking to maintain a continuous percept of the object with sophisticated processes [318].

Secondly, the paucity of sensory captors make perception somewhat of an ill-posed problem: the signal is insufficient to determine one unique interpretation. In visual processing, one can for instance know the distance and infer the size or reciprocally, know the size and infer the distance (the same phenomenon occurs with shape and angle) as shown in Figure 0.0.2.a. In auditory perception, the waveform hitting the eardrum can result from one unique source or else, from the combination of an unknown number of sources as shown in Figure 0.0.2.b.



(a) Visual ill-posed problem: From [333].



(b) Auditory ill-posed problem. From [296].

Figure 0.0.2: Perception is an ill-posed problem: the equation has too many unknowns and too few knowns to be resolved.

This indetermination cannot lead to an efficient behavior: if the cat cannot reliably evaluate the mouse's distance it will not be able to rightly calibrate its attacking jump. For ecological reasons, it is therefore necessary to resolve perceptual ambiguities.

Inferences

In the *Principles of Gestalt Psychology*, Koffka wrote : "The stimulus alone, no matter how defined, does not do justice to the perception" and further: "if things looked as they do because they are what they are, then perception would not contain in its very make-up a cognitive problem" [199]. From the

mere sensory input to the representation that one has of a complete scene, there are myriads of cognitive processes at play to fulfill the missing information. They can be compared to unconscious inferences [151] and the result of these inferences are a useful and stable representation of a perceptual object. Perceptual inferences occur beyond our conscious grasp, perhaps as a way to optimize speed processing [4]. Ambiguous stimuli are experimental objects designed to emphasize the inconclusive evidence carried by the sensory input and are used for the scrutiny of those fast and unconscious perceptual mechanisms.

While some perceptual organization principles may be implemented from a very early age [82], experience and knowledge may also be critical in the way we organize sensory input. Therefore, perceptual inferences may also result from the regularities we are exposed to throughout life. Bregman refers the principles derived from experience schema-based streaming principles and claims: "undoubtedly, there are learned rules that affect the perceptual organization of sound. I shall refer to them as schema-based streaming principles" and defines a schema as: "a mental representation of some regularities in our experience".

Context effects

Context effects refer to the processing of information external to the stimulus itself that contribute to process it in the most efficient way [354]. In visual perception a wide range of attributes such as motion, brightness, orientation, blur or higher-level structures like faces have been reported to be influenced by contextual information [335]. Figure 0.0.3 illustrates a simple visual context effect on size perception. In the left figure, the central red dot seems smaller than in the right one although they are identical. The surrounding information affects the relative sizes of the two objects.

Contextual information is particularly critical in auditory processing. Acoustic signals being inherently temporally defined, integrating information over

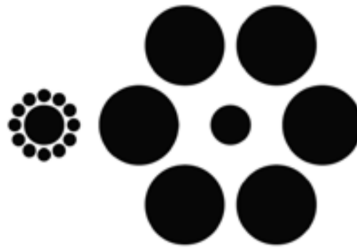


Figure 0.0.3: Context effect in visual perception. The central dot in the two figures seems to have a different size due to the surrounding dots although they are identical. From [378].

wide range of time scales is believed to be crucial in most complex auditory processing such as speech [183] and music perception [282].

Prior knowledge

A striking feature of human perception is that our subjective experience depends not only on sensory input but also on prior knowledge, acquired through experience and learning. Figure 0.0.4 illustrates how *a priori* knowledge biases visual perception (from [184]). The left dot seems convex while the right one seems concave. Why is it that an overwhelming majority of people will interpret the shadows this way? The reason lies in a very robust and long standing prior knowledge: object are lit by the sun from above.

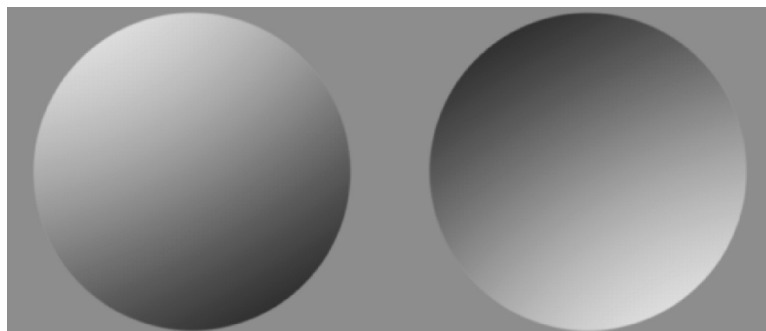


Figure 0.0.4: Prior knowledge effect in visual perception. The left dot seems convex while the second seems concave.

In the auditory modality, a phenomenon referred as the octave illusion is rooted in prior knowledge of listeners (demonstration: [Octave illusion](#)) [89, 246]. Two tones are presented continuously to both ears. Event A is a low tone presented to the left ear and a high tone presented to the right ear. Event B is the opposite: a low tone is presented to the right ear and a high tone is presented to the left ear. The low and high tones are pure tones separated by one octave. Events A and B alternated in a way that listeners were presented in each ear a succession of high and low tones and therefore, the high and low tones were always in opposite ears. However, listeners did not report this perception. They instead heard a single tone bouncing back and forth between the ears with an alternating pitch. Although it is still unclear whether this results from a suppression of the contralateral input or a fusion between the dichotic inputs [64], this subjective experience seems underpinned by the assumption, made by listeners, that a the signal is far more likely to come from a moving sound source, rather than two simultaneous sound sources arriving in each ear. The prior knowledge behind that effect can be explain in this words:

It is simply more probable in our world that an enduring object abruptly moved from one position to a nearby position than that one object suddenly ceased to exist and, at exactly the same instant, a separate but similar object just as suddenly materialized in another position.

From Shepard [344].

Prior knowledge could be defined as long-term statistical learning - the process by which the system encodes and learns statistical regularities in the world-. Empirical studies have suggested that statistical learning starts developing very early in life in both visual and auditory modalities [52, 320]. The precise mechanisms by which sensory information and prior knowledge are integrated remain unclear, with longstanding disagreement concerning whether integration is strictly feed-forward or whether higher-level knowledge influences sensory processing through feedback connections.

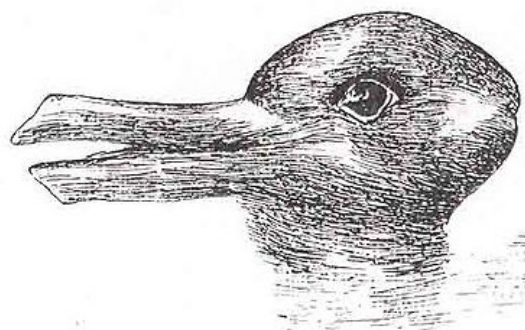
Sweet ambiguity

Ambiguity in laboratory

To cope with lacunary information, perceptual systems rely on various mechanisms such as context effects and prior knowledge to build a sensible and useful representation of the world. Despite its sophistication, the system may sometimes fail to accomplish this task. Perceptual illusions could, on first thought, be considered as failures. However being misled by our sensors is not necessarily problematic nor threatening. Constellations are based on the incorrect distances seen between the stars yet, although illusory, these objects are of great utility for sailors when searching their path at night.

Ambiguity, on the other hand shall be considered as a real perceptual defeat for it results in a perceptual *indecision* which could be problematic to guide one's actions. Consider the object is Figure 0.0.5. Is it a rabbit or a duck? Is it to eat or to fear? To chase or to repel? Those important questions (for, say, a predator allergic to duck meat) could not be answered as long as the indecision remains.

Figure 0.0.5: Jastrow's ambiguous figure of a duck-rabbit. Here can be seen either a rabbit figure or a duck figure. This example was then taken by Wittgenstein to lead his thoughts on general perception [406].



Extreme cases of ambiguous stimuli, such as those studied in laboratories have been frequently reported to induce a perceptual oscillation between two choices, making it impossible to settle for one with great certainty [230]. Luckily, such inconclusive sensory evidence is rarely encountered in real life and perceptual systems are rarely defeated, unless deliberately trapped. In laboratories, stimuli are purposely designed to carry maximal ambiguity, in order to allow scientists to observe the non sensory-driven mechanisms at play in perception. The Necker cube (initially described in Necker, 1832 [265]) is one of the most famous bistable stimuli in vision but ambiguity can also be studied in audition. Regarding the seemingly exclusivity between competing interpretations of a sensory input, Bregman argues that:

The exclusive allocation of evidence describes how these interpretations (vase/face) affect the edge that separate the vase from a face. When we see the face, the same edge is now located to the face. It is never allocated to both face and vase at the same time, but exclusively to one of them.

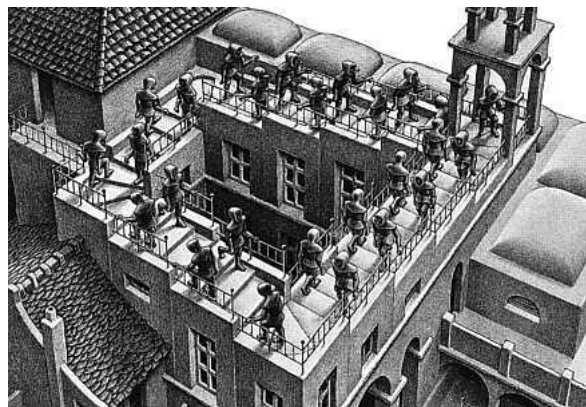
The claim is that in occurrence of ambiguity, the figures are alternating in the subjective experience and never present simultaneously. Arguably, exclusivity of competing interpretations is rooted in ecological reasons: it is more useful to settle on one percept at a time (a duck or a rabbit) rather than perceiving the midst of the two, which would correspond to nothing known (a duck/rabbit).

However, this implies that exclusivity may depend on particular ecological strategies endorsed by observers or listeners. Hence, it could be that in some cases, the midst of the two interpretations may be sensible, which could then not result in exclusive allocation of sensory information.

Ambiguity in art

Ambiguity have been extensively used for aesthetic purposes and artists of the 20th century were fascinated by bizarre perceptual effects. In music, ambiguous situations tend to generate greater emotions. In the lectures he gave at Harvard in the 1970s, Leonard Bernstein analyzed several examples of ambiguous excerpts in Western tonal music [220]. In the first bars of the slow movement of Mahler's 5th symphony, the tonal material is arranged such a way that the listener cannot decide if the main tonality is A minor or F major, as all the notes used by Mahler purposely belong to both tonalities. When the major F chord finally resolves the situation, the emotion aroused is even greater due to the perceptual indecision that the listener was deliberately left in.

Figure 0.0.6: Escher's stairs



More structural ambiguities fascinated M. C. Escher. The stairs displayed in Figure 0.0.6 constitute a perceptual paradox as they seem to go up indefinitely. This structure is similar to Ligeti's study for piano untitled *The evil stairs* which first bars are displayed in Figure 0.0.7. The left hand part is designed such a way that the listener is under the impression of a continuously rising pitch. The part highlighted in blue is what is responsible for this impression. This design is very similar to the Shepard tone paradigm that we will extensively evoke in this manuscript.

Figure 0.0.7: Ligeti's *Evil stairs*



Present research

This research is dedicated to study the perception of a case of an ambiguous auditory stimulus. The stimulus consists in an interval of Shepard tones. Shepard tones are complex harmonic tones which only contain octave-related components. When two Shepard tones are separated by a 6 semi-tones interval, i.e. a tritone, the perception of the pitch-shift is ambiguous: listeners will report with equal chances hearing an upward pitch-shift or a downward shift. This Ph.D work aimed at characterizing the processes underlying the perception of this ambiguous interval.

In Chapter 1, 2 and 3, a review of literature is conducted. Chapter 1 focuses on properties of perceptual ambiguity. Chapter 2 is dedicated to context effects. Chapter 3 reviews the main findings regarding the differences between musicians and non-musicians. The next three chapters present experimental results. Chapter 5 exposes three experiments conducted to characterize the temporal dynamics of a context effect biasing the perception of the ambiguous interval. Chapter 6 tackles an ongoing and seemingly unresolved question in the literature of ambiguous perception: do we perceive the ambiguity. Finally, Chapter 7 presents two experiments which further addressed

the question of the perceived ambiguity but here, it investigates the role of context effects.

PART I : Literature review

Chapter 1

Ambiguity in perception

Natural scenes, either visual or auditory, contain many ambiguities, due to inconclusive sensory input, that are typically unnoticed because they are resolved effectively in a fast and unconscious manner. Resolving perceptual ambiguities is therefore a common and frequent task, involved in most day-to-day perceptual processes. The specific mechanisms at play can be empirically explored using ambiguous stimuli, experimental objects designed to maximize the contradictory evidence. In this chapter, we will first briefly review the main paradigms that have been used to explore perceptual ambiguity, in the visual and auditory modalities. A second section will be dedicated to outline the general characteristics of perceptual ambiguity that emerge from distinct experimental procedures. Finally, we will review how inter-individual variabilities influence the behavioral outcome of perceptual ambiguity and what hypothesis have been proposed to account for these findings.

1.1 An hint of historical perspective

Perceptual ambiguity was first described in the visual modality by French physician Etienne François Dutour. In this study, published in 1760, he

observed that, when presented with a blue patch to one eye and a yellow patch to the other eye, the observer would not report seeing a mixture of the two colors but instead, his conscious perception would alternate between the two colors [103]. This phenomenon is now termed binocular rivalry and constitutes one of the main paradigms of perceptual ambiguity in the visual modality.

Auditory ambiguity have also drawn interest among scientists, yet to a lesser extent. The reasons for the prevalence of visual ambiguity lies into the fact that scene analysis has long been restricted to visual perception. Bregman, who inaugurated the field of auditory scene analysis with his seminal work, argues that auditory scene analysis entails as much complexity than visual scene analysis but may be at first more difficult to characterize [47]. Designing visual figures seems somewhat more straightforward, for visual items are better objects for introspective scrutiny. Auditory "objects" or stream, being transient in nature, do not offer the same grasp for phenomenological inquiries [315]. As a matter of fact, though it seems possible to sight what in a visual object is misleading, the same does not hold to the same extent in the auditory modality. This might explain the current relative paucity of auditory ambiguity paradigms.

1.2 The main paradigms

The next section will provide an overview of the main paradigms that have been used to explore the processes involved in resolving perceptual ambiguity.

1.2.1 Binocular rivalry

In binocular rivalry studies, the basics of the experimental procedure has not change much since its first description by Dutour in the 18th century. It

consists in presenting different images to both eyes, such as two gratings of different orientations, as illustrated in Figure 1.2.1. Something intriguing then happens: instead of a stable superimposed percept of the two images, observers report seeing an alternation between what is presented to each eye. It has become a very popular paradigm, for it allows to study the neural correlates of visual consciousness with great methodological flexibility: mere visual gratings are used to highlight encoding of simple properties, whereas more complex figures point to mechanisms occurring in higher-level of the visual pathway [43].

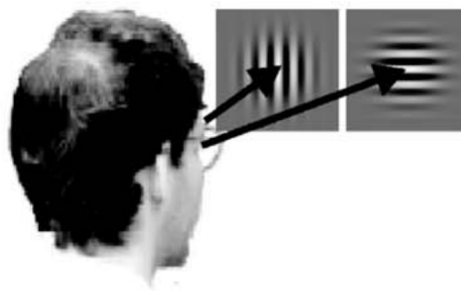


Figure 1.2.1: Procedure of binocular rivalry. Two different images are presented to both eyes. From [360].

1.2.2 Bistable visual figures

Bistable figures, such as those depicted in Figure 1.2.2 have also been extensively used to study perceptual ambiguity in the visual modality. It consists in presenting the same figure to both eyes and as in binocular rivalry, it is characterized by a switch between two alternative percepts. The term "bistable" refers to the alternation between two distinct percepts, and is now replaced by "multistable" that allows more than two alternative percepts to be perceived [86].

One of the earliest bistable figure studied is the Necker cube: a wire frame drawing of a cube in isometric perspective, which makes which depth perspective perceived ambiguous [265] (Figure 1.2.2, left panel). Other examples

of bistable figures, static and dynamical, are displayed in the other panels of Figure 1.2.2.



Figure 1.2.2: Illustrations of ambiguous stimuli. Left panel: Necker cube, from [360]. Middle panel: Rubin vase/face figure, from [371]. Right panel: schematic representation of an ambiguous apparent motion known as the « spinning wheel illusion », from [360].

1.2.3 Word repetition

The "word repetition" or "verbal transformation" paradigm was characterized as the "auditory analogue of the reversible figure" [393]. Indeed, repeating over and over the same word results in a puzzling perception: it entails a fluctuations of auditory percepts (illustration Figure 1.2.3).

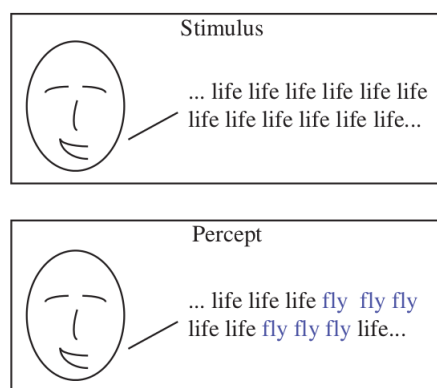


Figure 1.2.3: Word repetition paradigm. Figure by Agnès Leger from [334].

In an early study, twenty different phonetic sequences were repeated for three minutes of presentation during which listeners reported an average of 17.6 perceptual switches [394]. The switches occurred between different interpretations of the phonetic sequence, reported in Figure XX.

Ever since this first study was published, the word repetition paradigm has been used in several studies interested that took advantage of this tool to study the effect of vocabulary knowledge of listeners [395] and semantic content of repeated sequences [264] in the perceptual outcome of the effect. It was also implemented in fMRI (functional Magnetic Resonance Imagery) studies, in order to explore the contributions of non-sensory cortical structures in the dynamics of switches [202, 321] and EEG (ElectroEncephaloGram) studies to explore the increase of Gamma-bands activity (>30Hz) preceding perceptual switches [15], an activity known to parallel formation of perceptual objects [372].

Stimulus no. 2. Lame-duck: ...laimdukclaimduk... (58 forms). (9) duckling; (6) acclaimed; (5) claim-duck; (3) claim-de; (3) dark-claim; (3) claim-dock; (3) dock-lane; (3) ducklin; (2) came-dark; (2) claimed; (2) duck-lame; (2) I claimed.

Stimulus no. 3. Tress: ...trestrestres... (59 forms). (17) dress; (14) stress; (7) drest; (6) Esther; (5) best; (5) burst; (4) address; (4) stressed; (3) address; (3) best-drest; (3) spur; (3) stir; (3) tress; (2) a burst; (2) best-dress; (2) dress-best; (2) florist; (2) spurse; (2) spur-stir.

Figure 1.2.4: Results of a verbal transformation paradigm. List of the possible interpretations for the repeated word « Lame-duck » and « Tress ». From [393].

1.2.4 Ambiguous auditory streaming

Grouping and segregating sub-parts of the acoustic signal hitting the eardrum constitute the core of "auditory scene analysis" (ASA) [47]. The set of rules that characterizes ASA seem to be governed by a principle of continuity in the spectral domain (sounds whose frequencies correspond to the same "overtone series" tend to be grouped) [244, 227], the temporal domain (sounds with onset synchrony tend to be grouped) [399, 227] and spatialization (sounds with same source location tend to be grouped) [244, 101].

In successive sounds, ASA consists in allocating each event to a "voice" or "stream", a temporal object composed by the succession of sounds that are grouped together. Streaming abilities are believed to be crucial in auditory perception, as they appear very early in life [367, 405], severely impair people who lack them [129] and are not unique to humans [162, 114].

Experimentally, auditory streaming has been studied using sequences of tones A and B characterized by distinct frequencies, as illustrated in the top plot in Figure XX. The perceptual output corresponds to a triplet or galloping pattern "ABA-ABA", the grouped percept, or two separate streams "AAAA" and "BBBB" (the segregated percept) [360] (see Figure 1.2.5).

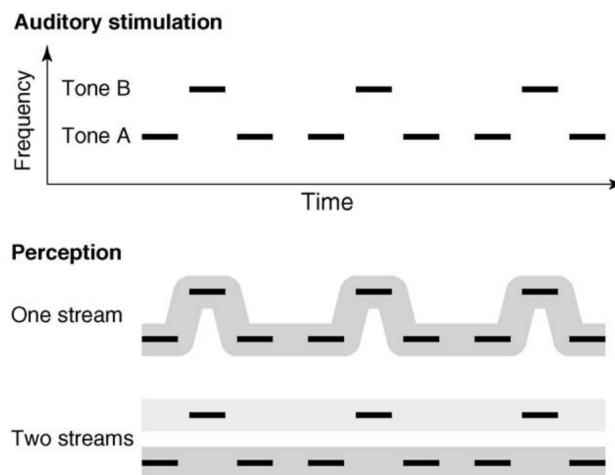


Figure 1.2.5: Illustration of the streaming paradigm. From [360].

Using this paradigm, Anstis & Saida demonstrated that the probability to segregate the signal in two streams increased with exposition duration [12]. The probability to hear two percepts is minimal at the onset of the signal and builds up over the course of presentation. This was interpreted to reflect an economy principle, according to which only the fewest number of sound sources should be considered to account for an auditory event [255].

Further research demonstrated that both temporal and spectral factors had an influence on the grouping of successive tones. Namely, small spectral

(ΔF) and temporal (inter-stimulus interval ISI) intervals between the tones induce a grouped percept [383]. Hence, a fission boundary was defined, as the combination of ΔF and ISI that would lead to maximally undetermined perceptual output.

Critically, when parameters are set to the fission boundary, listeners report either grouped or streamed sequences [295]. In this sense, auditory streaming under certain conditions -the "ambiguity region"-, is a case of bistable perception (Figure 1.2.6).

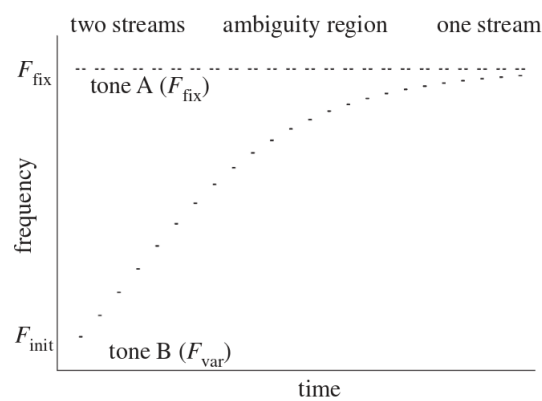


Figure 1.2.6: Ambiguity region in the streaming paradigm. From [257].

Denham et al. demonstrated that the "ambiguity region" correspond to a large number of ΔF and ISI combinations, for prolonged presentations [87]. It was observed that the first percept is maintained for a longer duration than subsequent percepts. It was argued that the mechanisms underlying both switches (initial and subsequent) is different. The first switch timing is determined by how long the system needs to consider the perceptual alternatives. Subsequent switches, on the other hand, arise from stochastic alternations whose dynamics rely on competitive interactions between coexistent interpretations (although a different interpretation has been proposed by Deike et al., see [81]).

Scientists have taken advantage of the methodological advantages carried by ambiguous auditory streaming to investigate the neural correlates of per-

ceptual ambiguity within the auditory modality. Indeed, different perceptual states are provoked by a constant stimulation and therefore ambiguity streaming afford the possibility to observe in real-time neural activities that parallel perceptual states [351, 144, 117] under various attentional conditions [74] and its effect on behavioral outcome and signal encoding [154, 369].

1.2.5 The Shepard paradigm

This Ph.D has focused on a perceptual ambiguity in the auditory modality that concerns pitch-shift direction¹. Pitch is a property of certain sounds which have the specificity to display periodic patterns over time and plays a crucial role in speech perception and music perception. To characterize pitch, scientists have felt the need to settle on a consensual definition formulated by the Acoustical Society of America, according to which it is "that property of a sound that enables it to be ordered on a scale going from low to high". Pitch is not defined by a single physical property (such as frequency), nor is it reduced to a physical concept: pitch is a psychological concept that arises from the interplay of various acoustical features (spectral envelope, loudness, frequency *etc.*) [78, 287].

Pitch possesses the curious property of having two dimensions. The fundamental frequency of a periodic sound (i.e. the lowest cadence at which it propagates) partially determines where it will be placed along the high-low scale. But pitch class -or chroma- is also an important trait of pitch perception and is strongly linked to music. Pitch classes of periodic sounds corresponds to the fact that, across the span of a one octave range, pitch can be labeled in different classes that will repeat for each octave cycle, as illustrated in Figure 1.2.7.

Therefore, pitch possesses a cyclic property according to which a sound corresponds to a pitch class or chroma that is invariant across octave transpositions. Strikingly, the octave cycle appears to be extremely broadly distributed

¹ It should be noted that "pitch-shift" here does not refer to the shift direction of the two fundamental frequencies of the two tones.

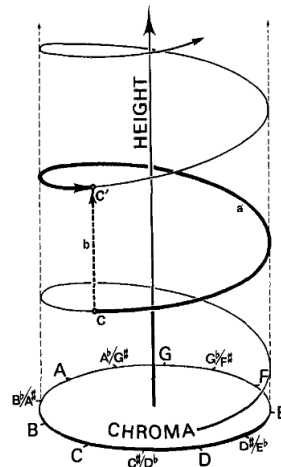


Figure 1.2.7: Helical representation of pitch along the dimension of height and chroma, or class. The highest in the helix a tone is, the highest the corresponding perceived pitch. The figure also captures the fact that two tones situated on the same vertical axis on the helix will have close pitch chroma. From [343].

across musical cultures [323, 267] which led many scientists to explore the cognitive and physiologic underpinning of "octave equivalence" [88, 253].

In the 1960s, Shepard designed octave-related tones, ever since referred as Shepard tones, with the intend to explore some of the psychophysical correlates of the octave cyclic representation of pitch [342]. Figure 1.2.8 displays a spectral representation of a Shepard tone.

Shepard noted that when two of these tones are separated by a 6 semi-tones (st) interval, also known as "tritone"² and which correspond to half of the octave (12 semi-tones), the resultant direction of the pitch-shift is ambiguous. The "ambiguity" here corresponds to the fact that some listeners report an upward shift, as if tone 1 was lower than tone 2, whilst for the exact same two tones, others will report the opposite direction, as if tone 1

² Amusingly, the 6st. interval, also called tritone, endorses a very special role in the History of music. Because of its extreme dissonance - to the point that its original sobriquet was *diabolus in musica*, "the Devil in music" - it has been cautiously avoided during all Ancient Greek and Medieval music, before it was increasingly used in Baroque and Classical periods to emphasize musical tension.

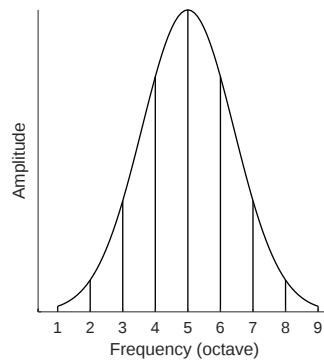


Figure 1.2.8: One Shepard tone. Each vertical line represent a frequency component whose amplitude is determined by a Gaussian envelope.

was higher than tone 2 [342, 53, 291]. The Shepard tritone hence constitutes a case of auditory ambiguity [77].

Interestingly, for other intervals, listeners' responses do not split up into two equiprobable directions ; they congruently settle on one or the other direction. The direction chosen follows what Shepard terms a "proximity principle" according to which frequency that are closer between the two tones will determine the directional perceptual outcome, a principal also observed in the visual modality [300]. The pitch-shift reported encompasses the smallest frequency region between components of the tones.

From that perspective, the ambiguity resultant of the 6 st interval is naturally explained: the equidistance of frequency components along a logarithmic frequency axis entails the absence of proximity bias. The direction is therefore unbiased toward one or the other directions by proximity cues [212], and the reported responses are accordingly evenly split between the two alternatives, as reported in visual bistability [260].

Deutsch et al. observed that listeners were consistent in the interval-shift they report for intervals with same frequencies [95], and that these idiosyncratic biases were robust to envelop modulations [132]. Long-term factors were proposed to be responsible for these idiosyncratic biases, such as voice-

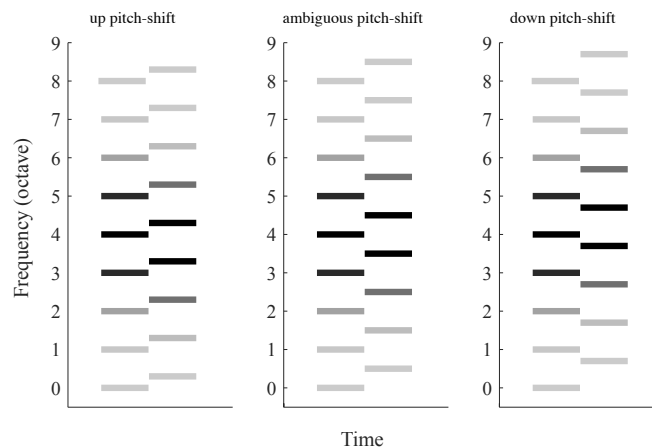


Figure 1.2.9: Proximity principle in perceived pitch-shift direction of Shepard tones intervals.

range [96] and early language exposure [94]. (however those findings have raised substantial controversy, see [308, 93]).

However, for certain frequencies (which are specific to each listeners), idiosyncratic biases do not seem to bias one particular direction and individual listeners' response is equally split between the "upward" and the "downward" interval-shift [95, 132]. This "50/50" response pattern, at the level of individual subjects, could result from two distinct subjective experiences: either from a clear pitch-shift direction that varies from trial to trial, or from an uncertainty between each pitch-shift directions that force listeners to respond "at chance". Although this particular question has been informally mentioned in several studies on Shepard tones [342, 311], (where anecdotal cases of listeners puzzled and confused by the ambiguity of their experience of the interval contrasted with other listeners being seemingly very confident in their perception of the pitch-shift), it has yet never been systematically explored.

1.3 General properties of perceptual ambiguity

Perceptual ambiguity can seemingly be observed in all modalities [61, 415], and display across-modalities interactions [322, 163]. Hence, it is a broad phenomenon, ubiquitous to perception in general [179]. Despite the diversity of paradigms used to study perceptual ambiguity, (use of stimuli from diverse modalities, steady or discrete presentations etc.), certain observations seem to be recurrent across paradigms [334]. In the next sections, we will expose the common features of perceptual ambiguity that can be outlined from diverse studies.

1.3.1 Similar behavioral observations

In and review confronting numerous studies on visual ambiguity, Leopold & Logothetis have identified three fundamental features of perceptual ambiguity in the visual modality [221]. The growing number of studies of perceptual ambiguity in other modalities [61, 414, 295] allow now to extend this inquiry beyond the visual domain. Which behavioral manifestations are recurrent in diverse studies on perceptual ambiguity?

It appears that the earliest noticed manifestations of perceptual ambiguity is that the steady presentation of an ambiguous stimuli induces the alternation of competing percepts, either in binocular rivalry [103], bistable visual figures [265], repeated phonetic sequences [393, 264] or streams of "ABA" tones [295]. Despite a undeniable role of attention and other top-down processes on a certain control over switch rate [270, 252, 86, 60, 366], the switches are ultimately experienced as "inevitable".

Interestingly, the temporal dynamics of switches display great similarities across different studies [5, 202, 295]. When measured over long period of steady presentation, dominance duration histograms (the time measured between two reported switches) describe what can be described as a Γ -like

distribution [5], or seemingly more accurately, as a lognormal distribution [416, 219, 209].

Critically, when comparing temporal dynamics of visual and auditory ambiguous stimuli, it was found that the distribution of dominance percepts did not significantly differ between the two modalities (see Figure 1.3.1) [295, 61].

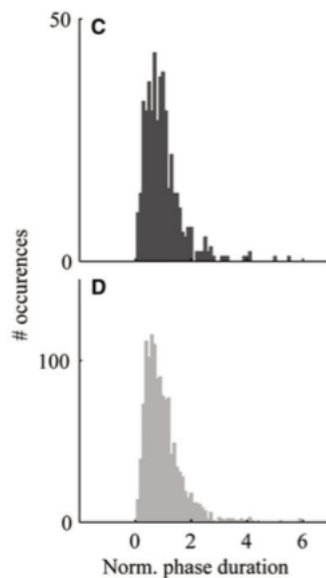


Figure 1.3.1: Histograms of dominance durations for the auditory (top) and visual (bottom) modalities. There is no significant difference between the two distributions. From [295].

1.3.2 Bottom-up and top-down integration

A sensory account of perceptual ambiguity

For binocular rivalry, it was hypothesized that the alternation between the two competing images resulted from a competition between sensory neurons [37]. Through reciprocal inhibition processes, the populations of neuron would compete against each other to alternatively reach dominance of firing

rates, provoking the switches between alternative percepts. Supporting this view, several studies demonstrated the strong implication of early post retinal [292, 283] and subcortical areas [149, 408] in the reversals of competing images.

In bistable figures, it was found that an eye movement shifting the focal center away from a bistable figure could reset the switch rate to its baseline [357]. Adaptation effects, whereby prior and long presentation of an unambiguous stimulus induce a contrastive context effect [11] were found to strongly support the sensory explanation of perceptual switches: the population of neurons encoding for the adapted percept being satiated from prior exposition, neurons encoding the alternative percept would display dominant activity [228].

In the auditory modality, similar hypothesis were proposed to explain the manifestations of verbal transformation. Warren & Gregory argue that alternations arise from a satiation of the representation of the first percept, which leads to a shift in the criterion applied to category boundaries between speech sounds, hence provoking a perceptual change [394]. The verbal transformation effect could arise from adaptation or fatigue-like phenomenon. Indeed early studies have established that repeated and sustained stimulation provokes adaptation of the neural detectors. Adaption of these detectors through repeated presentation may modulate stimulus encoding and provoke perceptual transformation.

Perceptual switches of alternative auditory streaming percepts were also found to be modulated by early sensory structures. An fMRI study suggest that transient BOLD signal is modulated in sub-cortical regions such as the thalamus and the supre-marginal gyrus [203]. Another fMRI study confirmed the involvement of such low-level structures : it reveals modulated activity related to perceptual switches within the inferior colliculus, one of the first post cochlear stages of auditory processing [324]. These recent results are of particular interest, for they suggest that areas purely dedicated to sensory processing are modulated during bistable perception of a constant

input. As pointed out by Gutschalk et al., further studies will have to reveal to which extent selective attentional gain ([314], for a recent review see [67]) is responsible for these puzzling findings [143].

Non-sensory structures modulated by perceptual switches

However, the implication of higher-level structures, including non-sensory areas, was also reported in different studies on perceptual ambiguity [186, 361, 48, 191]. In visual bistability, fMRI fluctuations paralleling binocular rivalry were found in cortical sensory areas, such as V3, V4 and MT (for a review see [360]). It was found that suppressed stimuli still elicit activity in areas not dedicated to sensory processing [403, 113]. Wang et al. demonstrated that the implication of these high-order brain regions was more pronounced in bistable perception than in classical, non-ambiguous perception [391].

In verbal transformation, implication of non-sensory structures in perceptual reversals were found [202, 321]. Kondo et al. conducted an fMRI study in which the word "banana" was repeated 265 times during a 90s session, while participants reported verbal transformation by pressing a button [202]. An event-related fMRI analysis revealed that the left inferior frontal cortex (IFC), anterior cingulate cortex (ACC), and the left prefrontal cortex were activated during perceptual transitions. Intra-subjects correlations revealed that a higher reversal rate was paralleled by an increased activation of the left IFC and a decrease activation of left ACC (Figure 1.3.2), leading them to conclude: "the left IFC activity facilitates changes in verbal forms, and the ACC activity suppresses them".

In ambiguous streaming of successive tones, it was suggested that some non-sensory structures were implicated in the grouping of auditory input [74, 155], echoing findings in vision [341].

Another line of evidence consisted in showing that lesions in certain non-sensory areas could impact the behavior of observers when confronted to

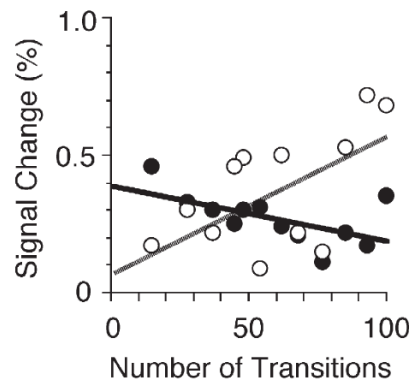


Figure 1.3.2: Intra-individual regression analysis. Each circle corresponds to an individual point of signal intensity as a function of the number of perceptual transitions in the left anterior cingulate cortex (ACC) (closed circles) and left inferior frontal cortex (IFC) (open circles). From [202].

ambiguous stimuli. Ricci & Blundo reported that patients suffering frontal lesions -known to be the siege of attentional processes, for a review see Corbetta et al. [71]- had great difficulty perceiving the two interpretations of the ambiguous figures as compared to control subjects and subjects suffering more posterior lesions [313]. This was confirmed by another study implicating subjects who had a focal cortical excision (as a treatment to their epileptic condition). The authors observed the same inability to perceive more than one interpretation in subjects whose small part of the frontal cortex had been removed as compared to control ones [245], hence implying a key role of frontal regions in the reversal process.

Attention and prior knowledge effects

Supporting the view of a strong implication of top-down processes, several studies have emphasized the role of attention in modulating the dynamics of switches. It was shown that exogenous attention is very efficient at

determining which of the two percepts will reach dominance on the first presentation [270, 252] or which will be maintained longer [214, 66]. Volition control, for instance, was found to have a strong impact on the rate of switches [226, 366]. Toppino demonstrated that the effect of volition was independent of the focal position with respect to the figure, suggesting that cognitive control can overcome the effect of sensory adaptation [380].

Semantic knowledge was also reported to trigger perceptual switches. A study involving high school students as observers demonstrated that the rate of reversals drastically changed once the observers were told about the ambiguous nature of the figure [134]. This was confirmed by a later study led by Rock *et al.* on children from three to four years of age. Unlike previously observed with adults, none of the children spontaneously reversed the figures after 60s of presentation [316]. A possible interpretation could be that children of this age have been very little exposed to such figures as compared to adults. Yet, Rock argues that findings in adults reflect methodological flaws whereby subjects are biased by experimenters' knowledge of the reversal possibility.

A recent study replicated Rock's finding on ten years old children whose reversal rates were compared to control adults' ones [106]. It was found that adults displayed a much higher rate of perceptual switches than children, paralleled by a fronto-parietal activity that was not observed in children's recordings. However, the results of this study suggest that children have less perceptual switches due to immature frontal structures [125, 133] rather than different prior knowledge.

A clearer effect of semantic content was found in verbal transformation studies. Repetition of meaningful words (e.g. dollar, seven) prolongs the occurrence of verbal transformation as compared to the repetition of meaningless words (e.g. rollad, neves) [264]. Moreover, children of six to ten year of age report more non-sense words than adults, an effect likely due to the difference in vocabulary storage in childhood [395]. In ambiguous auditory streaming, after they were instructed what were all possible interpretations,

listeners experienced more spontaneous switches than before training [86].

High order context effects, such as semantic context, may also bias observers towards one or the other interpretation. Bruner & Minturn demonstrated that the image displayed in Figure 1.3.3 could be seen either as the number "13" or the letter "B" depending on whether letters or numbers surrounded it [51].

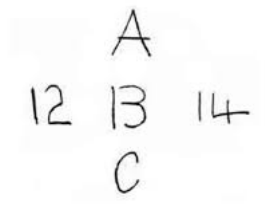


Figure 1.3.3: The central item can be perceived as the number "13" or the letter "B" depending on which informations is surrounding it. From [51].

Integrative model

In various paradigms concerned with perceptual ambiguity, empirical evidences that favored a sensory account of the dynamics observed in perceptual ambiguity, both in visual [292, 283, 149, 408] and auditory [203, 324, 77] modality were collected. However, several elements, such as behavioral observations of the role played by attention [226, 270, 252] and prior knowledge [134, 264, 316, 395], neuropsychology studies showing specific impairment in the perception of ambiguous stimuli [71, 313, 245], and evidences from imaging studies of the implication of non-sensory structures in visual studies [186, 361, 48, 191, 403] and auditory studies [202, 321, 154, 74]. Altogether, these findings advocate for an integrative model where bottom-up information encoded in sensory areas would be modulated by top-down processes, the result of which determines the contents of awareness [230, 205, 50].

In binocular rivalry, Tong et al. proposed an integrative modal in which

excitatory feedback projections from high-level structures modulate the suppressive activity of early visual areas (Figure 1.3.4) [379].

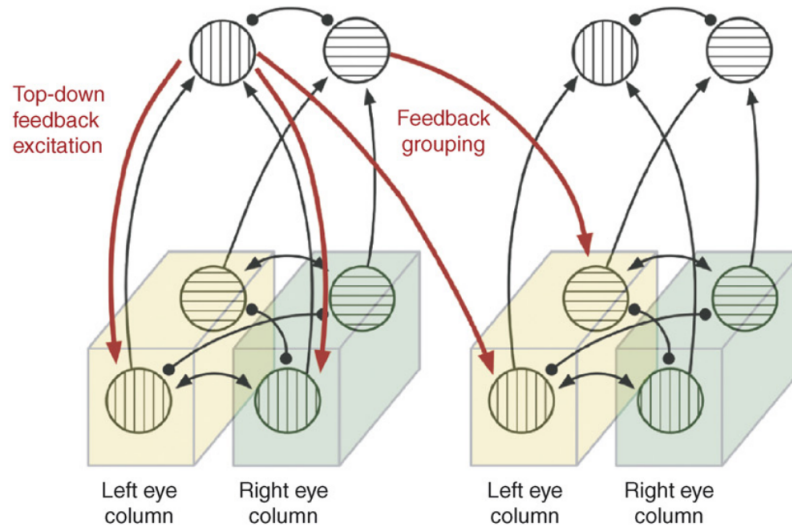


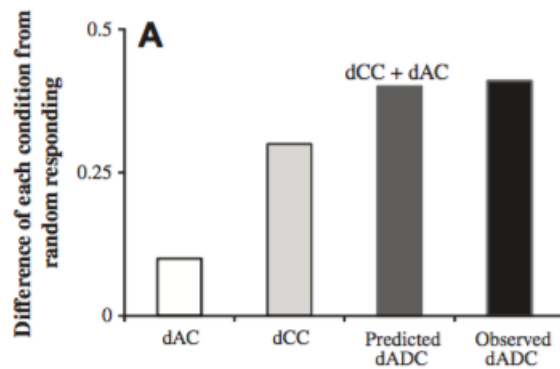
Figure 1.3.4: Integrative modal of lateral inhibition and excitatory feedback projections. From [379]

In visual bistability, an elegant study conducted by Intaite et al. was conducted with the aim of disentangling the respective contributions of bottom-up and top-down processes on the perception of ambiguous figure [165]. The ambiguous stimulus consisted in a two-dimensional drawing of a Necker cube-like figure composed of five overlapping squares (left panel of Figure 1.3.5). The design consisted in observing the effects, separately and jointly, of two types of context effects on the ambiguous stimulus. The context stimulus was the same in both cases, but its type of presentation was supposed to induce different effect on the ambiguous object. It was an two-dimensional drawing of five overlapping squares with not ambiguous depth (top panel of Figure 1.3.5). In the adaptation condition (AC), it was supposed to induce a contrastive effect [228]. In the simultaneous context condition (CC), it was supposed to induce an attractive effect, where the perceived interpretation is cue-compatible with the context figure [136, 389].

They had four conditions, two in which they tested for each effect separately and two in which they tested the interaction of either congruent or incongru-



- (a) Experimental design: prior presentation and simultaneous presentation effects of an ambiguous figure of overlapping Necker cubes.



- (b) Results: The differences from random responding of adaptation (dAC), first bar and context (dCC), second bar. Third and fourth bar corresponds to the predicted (if both effect are additive) and observed summed effects in the congruent condition.

Figure 1.3.5: Experimental design and results from Intaite et al. [165].

ent predicted output. The results (bottom panel of Figure XX) display the results normalized by the random level for each effect separately (two first bars) and observed added effects for the congruent condition (fourth bar). These results demonstrate that the perception of the ambiguous stimulus was modulated by both types of effects, adaptation from prolonged prior exposition (AC) [229] and simultaneous unambiguous stimuli (CC) [136, 389]. When biasing congruently the perceptual output of the ambiguous square, both effect were additive, thus advocating for additivity of top-down and bottom-up processes in the perceptual ambiguity [205, 336].

1.3.3 The special case of exclusivity

In their influential paper, Leopold & Logothetis stated that exclusivity was one of the three fundamental hallmarks of perceptual ambiguity [221, 295]. However, recent findings reported that observers may report mixture percepts associated with greater ambiguity, suggesting that the hallmark of exclusivity might more controversial than primarily believed [295, 86, 188, 251, 348].

Exclusivity corresponds to the fact that in perceptual ambiguity, alternative interpretations are exclusive to each other, i.e. are not perceived simultaneously. Although collecting mixture percepts was done in an early binocular rivalry study [157], later studies took exclusivity for granted, as they only allowed participants to report one or the other percept. Yet, recent studies sought to determine whether exclusivity was such a stable feature, as they allowed participants the possibility to report undetermined or mixture percepts, as illustrated in Figure 1.3.6 [295, 86, 188, 251, 348].

Behavioral studies in which more than the two main percepts could be reported by participants have somewhat gathered contradictory findings on this matter. In some of them, mixture percepts were very rare [295], thus suggesting that the possible perceptual interpretations were exclusive to each other, whereas other have reported that mixture percepts were experienced

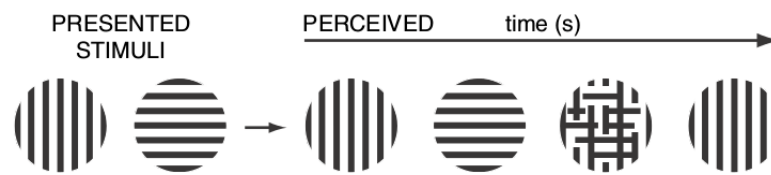


Figure 1.3.6: Results of Miller et al. Example of mixture percept report in binocular rivalry procedure. From [251].

quite often [86, 188]. Whether this discrepancy reflect mere procedure differences from studies to studies or is indicative of distinct neural and cognitive phenomenon underlying the presence or absence of exclusivity between the possible percepts is yet an open question and will need further investigation.

The neural bases of exclusivity

From a neuronal point of view, exclusivity may be accounted for by mutual inhibition mechanisms. The more a neural representation is inhibited, the more the corresponding percept is suppressed, hence yielding to exclusivity. Findings supporting this view were found: exclusivity was found to depend on stimulus strength in binocular rivalry [157]. In a recent study, very long exposition of binocular rivalry using oriented gratings were presented to observers while they reported either one or the other of the exclusive percepts or mixed percepts (see left panel of Figure XX). The results revealed that the proportion of exclusive percepts decayed over time, as illustrated in the right panel of Figure 1.3.7.

To account for these intriguing results, the author argue that exclusivity arise from the counterpart of the famous "neurons that fire together wire together" proposed by Hebb [150], and according to which neurons that frequently fire simultaneously tend to strengthen their synaptic connections through long-term potentiation [38]. Reciprocally, neurons that do not tend to fire together should unwire, a phenomenon referred as anti-Hebbian plasticity [404]. After prolonged exposition that first elicited exclusive percepts, the system tends

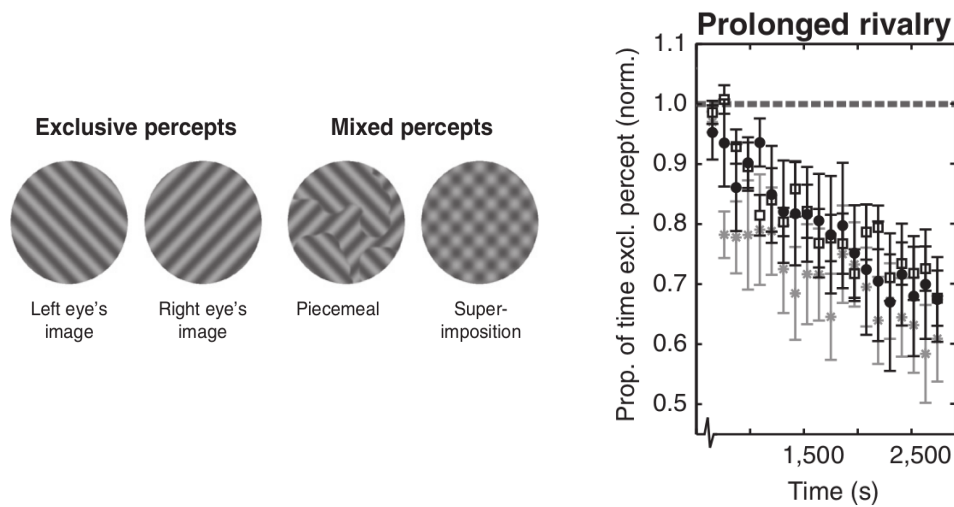


Figure 1.3.7: Results of Klink et al. Left panel displays the possible alternative percepts that observers could report. The right panel corresponds to the proportion of exclusive percepts over time averaged across participants and computed for 100s stimulation. From [188].

to strengthen new bounds between neurons that suddenly started to fire together, as a coping mechanism to the new environment, thus inducing non exclusive percepts to build up (see also [288]).

Vagueness: a case of ambiguity without exclusivity?

Vagueness is a concept that originally belong to logic and philosophy but has recently been experimentally invested (for a recent review, see [7]). Vagueness is intrinsically coupled with the phenomenon of soritical series. A soritical series is a composition of gradually changing items so that each extreme of the series clearly belong to distinct categories and middle items constitute borderline cases. For instance, a series of color shades that gradually changes color so that the first item clearly corresponds to a green category and the last item clearly belong to a blue category (as illustrated in Figure 1.3.8), is a soritical series.



Figure 1.3.8: Example of soritical series of color shades gradually changing from a clear green to a clear blue. From [105].

Vagueness corresponds to the fact that no definite line can be drawn at one point of the series that would fairly mark the transition from one category to another: the mark point of the transition always appears to be arbitrary to some extent. Raffman makes the claim that border line cases in a soritical series constitute instances of ambiguous stimuli, for their perception entail uncertainty regarding which category they belong to [298].

Assigning stimuli to a certain category is a task named categorical perception [148]. A recent fMRI study was conducted to investigate the neural bases of categorical perception. Observers were presented vertical lines of different length that they had to assign to the category "short" or "long" [140]. This judgment had to be made according to two different criterion lengths. The criterion length were not shown directly but learned implicitly through feedback. Prior to the categorical judgment task, participants were informed of which of the two criterion they had to compare the test stimulus by a color code. Probe lines that were close to the criterion may be considered as borderline cases of a soritical series whose two extremes would be "short" and "long".

An psychophysical uncertainty function was computed for each participant ($n = 10$) so that uncertainty ranged from 0 to 1. Figure 1.3.9 displays the correlation between uncertainty and response time averaged across participants.

These results reveal that uncertainty of categorical judgment increases the time that participants took to report their response. This effect was quite large, as a typical participant's response time ranged from 0.5s to 2.5s.

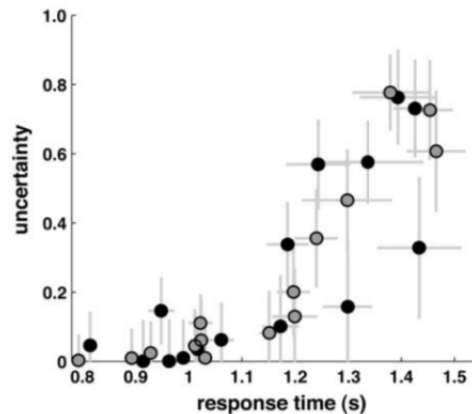


Figure 1.3.9: Results from Grinband et al.. Linear regression between uncertainty and response time averaged across participants ($n=10$), $R=0.342$, (error=SD). From [140].

Therefore, consistently with Raffman's view, the categorical judgment of borderline cases entail uncertainty toward which category they belong to which results in longer response times. It might be interesting to put these results in perspective with the diffusion model proposed by Ratcliff [302]. According to this model, response latency is a function of the distance between the stimulus and the decision criterion [111].

In this respect, vague stimuli are very similarly to ambiguous stimuli. However, one could argue that a substantial difference between the two lies in the fact that vague stimuli entail a conscious representation of the two categories and of the internal criteria separating them. Categorical judgments -typically involved in the processing of vague stimuli- then consist in comparing a particular stimulus to this internal criteria. The proximity of the stimulus to the internal cutoff between the two categories makes the judgment more difficult and this can be observed through increased response times. From that perspective, exclusivity would hardly be a hallmark of perception of vague stimuli.

In a recent review on multistability, Schwartz et al. develop the argument that vague and ambiguous stimuli differ in nature because they do not en-

tail the same binding mechanisms: a distinction is made between "boundary stimuli" which have features close to a boundary between two perceptual categories along a perceptual continuum and "ambiguous binding stimuli" which induce ambiguity regarding the whole binding organization. Because of this distinction in feature organizations, the later typically involve multi-stability while the former do not. Yet, as pointed out by the authors, this discussion has received theoretical consideration (see [104]) but yet, it still needs to be addressed experimentally [334].

Response times: a marker of exclusivity?

If response times increase as a function of uncertainty in categorical judgment where vague perception occurs, shall we observe the same effect in ambiguous perception as the degree of ambiguity varies? This question has been explicitly addressed by Takei & Nishida in a study in which the degree of ambiguity of two visual stimuli was varied in order to observe the effect on response latency [371]. A rotative cylinder, ambiguous with regard to the 3D rotation direction (see Figure 1.3.10) was used in a first experiment.

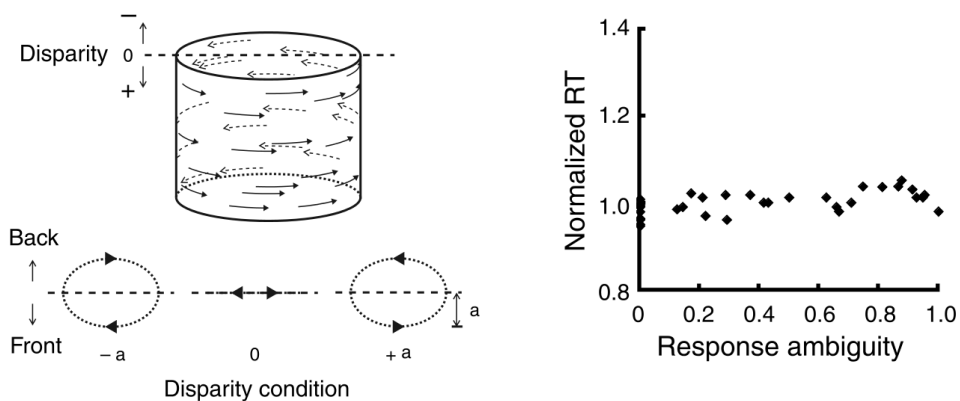


Figure 1.3.10: Stimulus used in the first experiment in Takei & Nishida. From [371].

A disparity cue was used to control for the magnitude of this perceptual ambiguity. The disparity cue was varied so to bias observers' perception toward

one or the other direction and participants were tested on seven disparity conditions. Observers were instructed to report, as quickly as possible, the perceived rotational direction. Results from response latencies are displayed in the right panel of Figure XX. The normalized reaction times are displayed as a function of the response ambiguity.

A second experiment was conducted, using the Rubin vase/face figure, as illustrated in the top panel of Figure XX. The contrast of the image was varied in five contrast conditions in order to bias observers' perception towards one or the other interpretation and as the previous experiment, observers were asked to report their percept as quickly as possible. The response times as a function of response ambiguity for this task are plotted in the bottom panel of Figure 1.3.11.

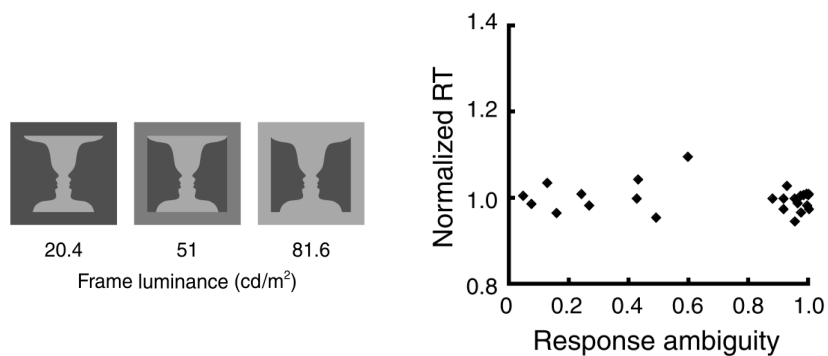


Figure 1.3.11: Normalized reaction times (RT) plotted against response ambiguity. Each point corresponds to a participant RT and response ambiguity for each five disparity conditions. From [371].

The two experiments consistently demonstrated that response latency did not significantly increase with response ambiguity. Besides, these results echoed precedent findings from an experiment led by the same team, using ambiguous motion quartet [370]. These results comfort the independent race model according to which evidences for each perceptual interpretations are accumulated independently and as a consequence, do not entail an increase in response latency. In favor of this view, one could argue that -as discussed

earlier-, ambiguity is a largely distributed phenomenon whose processes are most likely at play in "natural perception". In this regard, the perceptual system should take fast decisions in spite of contradictory elements.

Interestingly, findings in contradiction with those of Takei & Nishida were recently found in a binocular rivalry paradigm [168]. Their experiment consisted in presenting moving dots in two monocular presentations, dots from both images moving either in the same direction (unambiguous), or in opposite (ambiguous) directions (Figure 1.3.12).

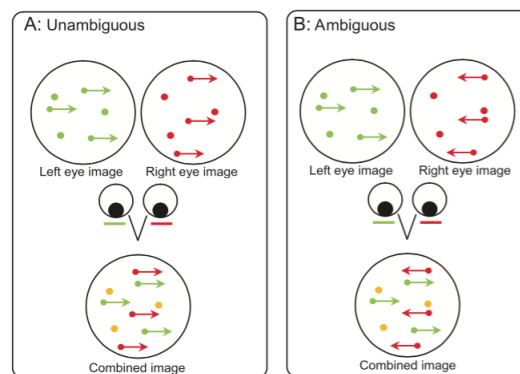


Figure 1.3.12: Schematic illustration of the two conditions (unambiguous and ambiguous) in experimental method. From [168].

Their result showed that in the ambiguous case, the report of the direction perceived took overall longer time than when the dots presented to the two eyes were moving in opposite directions (see Figure 1.3.13). A control condition checked the change in reaction time was not explained by a transparent motion percept in the ambiguous conditions. They hence interpret the slowing down of the decision process as the result of the sub-conscious presence of the rivaling stimulus, in accordance with other results showing an impact of perceptually suppressed stimuli [9].

The discrepancy with Takei & Nishida study could be rooted in the important methodological differences between the two studies. In Takei and Nishida, the ambiguity arose from the competition between two interpretations of the same visual stimulus [371]. Their interpretation of the absence of effect on

response time was rooted in ecological reasons: since resolving perceptual ambiguities is probably a frequent task in general perception [179], it is sensible that the choice of one interpretation over the other is made in a fast and unconscious manner. In Kalisvaart et al., the ambiguity was created by opposite movements of dots in both eyes [168], a situation that is not encountered in most day-to-day perceptual scenes.

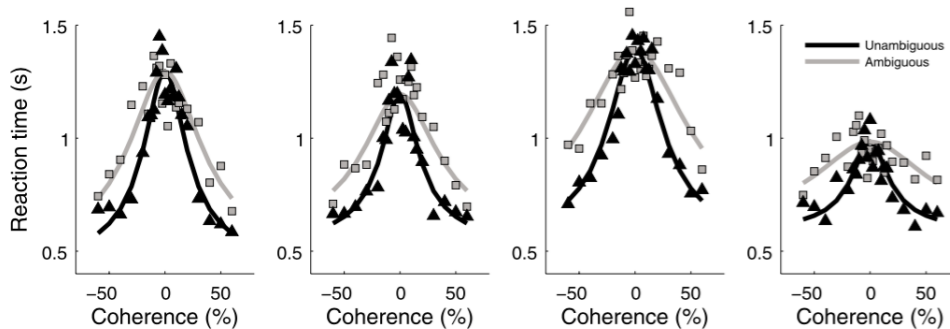


Figure 1.3.13: Mean reaction times results. Each plot corresponds to individual data. Black triangles: Unambiguous stimuli. Gray squares: Ambiguous stimuli. Arrows marked L and R underneath the horizontal axis (top) show the direction of motion in the left and right eyes, respectively, in the ambiguous condition. Solid lines show the model fit. From [168].

To conclude on this section, response latencies have so far been seldom measured in perceptual ambiguity studies. The reason may lie in the fact that methodological reasons: spontaneous reports of perceptual switches from subjects do not allow to compute response times, for the time at which the endogenous change occurs is challenging to time [365]. This could also be explained by theoretical reasons: exclusivity has long been considered as a hallmark of perceptual ambiguity [221] and it is only fairly recently that this question started receiving increasing attention [295, 86, 188, 251]. Hence, using response times measures as a marker of exclusive perception may become more and more popular.

The very few elements gathered so far point to puzzling findings [371, 168]. While Takei & Nishida did not observe an increase of response times with

increasing stimulus ambiguity, Kalisvaart et al. report an increased response latency associated with an exclusive percept (observers were not conscious of opposite motions in each eye). This suggests that response times alone could not be sufficient to determine the presence of exclusivity in perceptual ambiguity. The coupling of response times and confidence ratings could therefore be used in another setting as a new tool to address the detect the presence of exclusive perception for ambiguous stimuli [284].

Response time, marker of exclusivity? To further explore this hypothesis, it would be interesting to observe response time modulations as the presence of mixture percepts builds up [188, 288]. This suggests that response times alone could not be sufficient to determine the presence of exclusivity in perceptual ambiguity. The coupling of response times and confidence ratings [387, 303, 135] could therefore be used in a novel setting, as a powerful combination to detect the presence of exclusive perception for ambiguous stimuli [284].

1.4 Ambiguity across individuals: inter-subject variability

In a last section of this chapter, we will review the main findings gathered on the inter-subject variabilities in perceptual ambiguity. Inter-individual differences are typically treated as a source of noise in experimental results and possible explanations for those differences are discarded through averaging data [171]. However, inter-individual variabilities may also be treated as a source of information for the exploration of cognitive mechanisms, such as auditory perception [243, 401]. In perceptual ambiguity, great inter-individual variabilities had been reported since the 1960s. Aajdes et al. have reported great inter-individual variability -up to an order of magnitude- in the perceptual rate of perceptual switches in binocular rivalry [1]. Inter-individual variabilities have thus been used as a tool to tackle particular questions of

perceptual ambiguity, such as the types of structures involved in perceptual switches [202], or the common origins of processes across different modalities [295, 204].

Recently, an upsurge of interest has drawn back attention on individual differences *per se* [187]. Perceptual ambiguity arise from the interplay of different levels of processing, including sensory and non sensory structures. To date, the architecture underlying such interplays remains partially unknown. Inter-individual variabilities may be a suitable way to explore further this architecture for they reveal what are the neural basis of variability in perceptual ambiguity [284].

1.4.1 Behavioral evidences for individual differences

While using inter-individual differences as a tool for addressing other questions on perceptual ambiguity, several findings consistently demonstrated great variability in behavior.

Pressnitzer & Hupé investigated whether temporal dynamics of bistable perception could be compared across modalities by using inter-individual correlations within and across modalities [295]. They found individual consistency in perceptual switches rates within modalities but not across modalities. In a word repetition experiment, Kondo & Kashino used intra-subjects correlations to demonstrate that a higher reversal rate was paralleled by an increased activation of the left inferior frontal cortex and a decrease activation of left anterior cingulate cortex (Figure 1.4.1) [202].

In a recent study, Denham *et al.* tackled the question of stability of inter-individual variability over long period of time [86]. In a first experiment, they presented "ABA" triplets to listeners and trained them to report all possible perceptual interpretations. Participants had to their disposal a response interface composed of six possible perceptual interpretations that they were invited to use whenever they switched from one percept to another. First

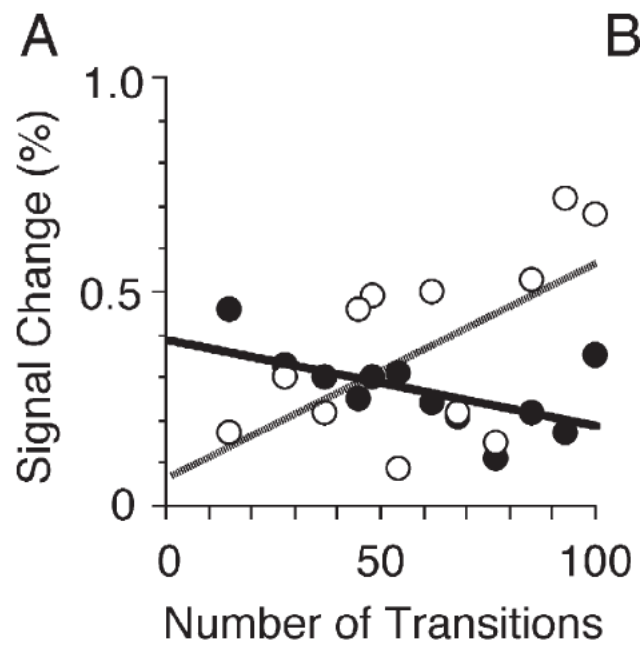


Figure 1.4.1: Intra-individual regression analysis. Each circle corresponds to an individual point of signal intensity as a function of the number of perceptual transitions in the left anterior cingulate cortex (ACC) (closed circles) and left inferior frontal cortex (IFC) (open circles). From [202].

it appeared that, after training, participants used more than the classical segregated vs. integrated switch, as they experienced other patterns. The experiment took place over the course of one month during which six participants took part in seven sessions.

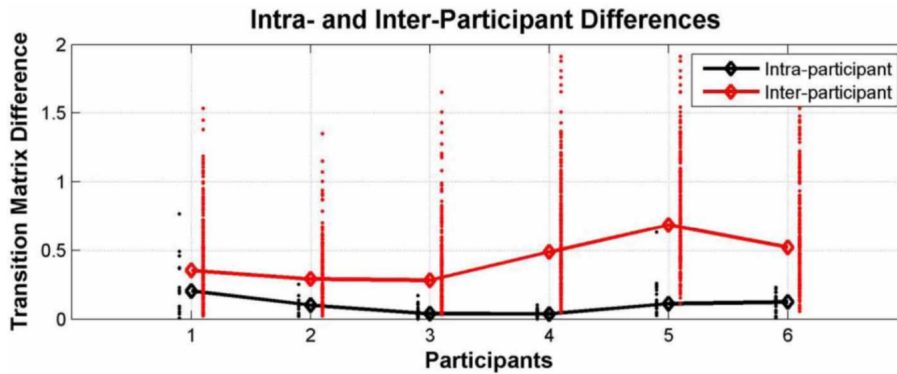


Figure 1.4.2: Results from experiment 1 in Denham et al.. Differences between transition matrices from the same participant at different sessions (black dots; median—black diamonds, line), differences between the transition matrices of each participant and those of all other participants (red dots; median—red diamonds, line). From [86].

Transition dynamics were computed for all participants and allowed an intra- and inter-participant variability analysis across all the sessions. To do so, a transition matrix was constructed, by counting the number of occurrences of each transition for each participant and each sessions. Figure 1.4.2 displays the inter- and intra-individual perceptual switching behavior. Differences between transition matrices from the same participant at different sessions are represented by black marks. Differences between the transition matrices of each participant and those of other participants are represented by red marks. This analysis revealed that participants' behavior is idiosyncratic: the switching rate of a given participant is more similar to its own switching rate in different sessions, than to other participants' switching rates.

A second experiment was run with the same stimuli, procedures and participants, one year later. In order to test for stability of idiosyncratic traits

over a long period of time. This experiment revealed that differences in individual switching bias remained consistent over a long period of time. The authors concluded that such stable perceptual traits possibly stemmed from genetic differences between listeners.

1.4.2 Genetic explanations

Kondo et al. gathered the first evidences suggesting that switches may be modulated by genetic factors [204]. They assessed that variability in perceptual rates of switching among individuals were correlated to genetic factors. It was also suggested that participants suffering from a bipolar disorder -known to be highly heritable- exhibit slower perceptual rates than control participants [250].

Miller et al. took this inquiry a step further as they conducted a study using the twin method on a large sample of twin pairs. The authors recruited two different types of twins: monozygotic (MZ) twins ($n = 128$ pairs) who are genetically identical, and dizygotic (DZ) twins ($n = 220$ pairs), who share roughly half of their genes. The rationale behind twin studies rests on the differential of common genetic information in different types of twin pairs, whilst environmental factors remain similar for both pairs. Therefore, it offers an opportunity to isolate purely genetic factors underlying different behaviors. In this case, the behavior is perceptual switch rate in binocular rivalry, known to show great inter-individual variability [1].

All participants completed several sessions of binocular rivalry. They were presented gratings of different orientations in each eye and were asked to report whether they were seeing one or the other, or a mix of both. A perceptual switch rate was computed for each of them. A correlation coefficient could then be computed for the two types of twins, MZ and DZ. Figure XX displays the results of comparison of different behavioral measures correlation between different types of twin pairs.

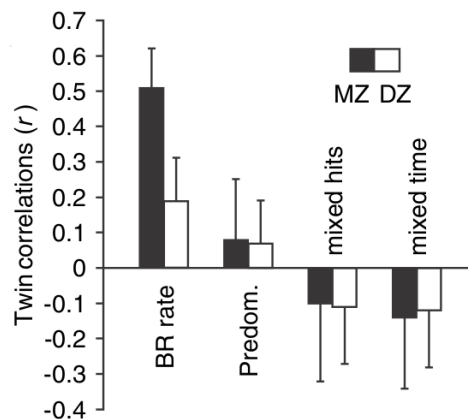


Figure 1.4.3: Results of Miller et al.. Correlation coefficients are compared between MZ (black bars) and DZ (white bars) twin pairs. The pairs of bars correspond to perceptual switch rate, predominance time, number of mixed percept reports and time spent on mixed percept respectively (error bars indicate 95% CI). From [251].

The first two bars correspond to perceptual switch rate (BR rate) correlations within the MZ and DZ twin pairs. Subsequent bars correspond to comparison of predominance -a measure of the ratio between the two possible percepts-, and measures relative to mixed percepts. It appears that only BR rate displayed significant difference between the two type of twins. In contrast, none of the other measures differed between the two populations. In other words, being MZ twins make individuals much more likely to experience the same switch rate than being DZ twins. Considering that MZ twins have -nearly- the same genes, these results demonstrate that BR rate in binocular rivalry is highly heritable. The authors could assess that 52% of the variability in spontaneous switches can be accounted for by genetic factors. To my knowledge, this study represents the first firm evidence that behavioral peculiarities in the experience of ambiguous figures are strongly determined by genetic factors.

1.4.3 Neurophysiological explanations

Another way to explore the substrate of individual variability in behavior is to relate it to individual variability in brain structure using morphometric analysis [187]. Kanai et al. conducted two MRI studies in which they tackled this question [169, 170]. An ambiguous motion was presented to participants while they reported perceptual switches. Consistently with previous findings, great inter-individual variabilities were found in percept durations and switch rates. These idiosyncratic features were correlated with morphometric measures using MRI. The analysis revealed a significant link between cortical thickness and percept durations in bilateral superior parietal lobule.

A further study from the same authors refined which were the involved brain structures and demonstrated that bilateral anterior SPL were critically involved in perceptual rates variability. A positive correlation between grey matter density in these regions and percept durations could be plotted, as illustrated in Figure 1.4.4. It suggests that the thicker the cortex in these regions, the faster the switch rate of the brain's owner.

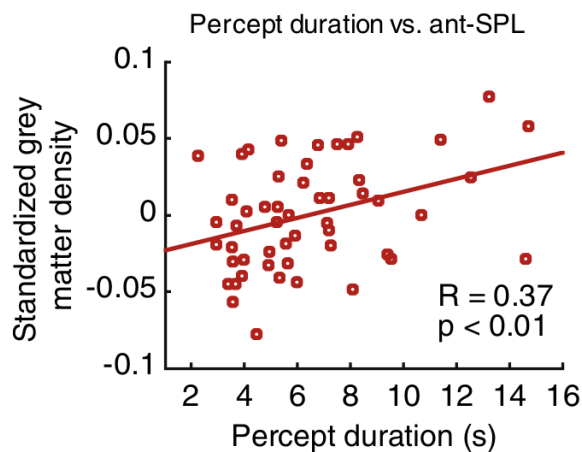


Figure 1.4.4: Results of Kanai et al.. Individual standardized grey matter density in the central voxel of the region of interest are plotted against percept durations. Each dot corresponds to a participant (n=52). From [170].

Finally, another line of evidence has recently demonstrated interesting results. Individual differences in perceptual dominance preferences in an ambiguous motion have been related to microstructural interhemispheric connections [127].

These results afford very interesting insights on the neural substrate underlying individual variabilities in behavioral measures of perceptual ambiguity. Nonetheless, it remains unclear whether these structural differences arise from genetic factors or whether they result from experience and plasticity. Indeed, a considerable degree of plasticity is retained throughout life and may possibly change the neural structures engaged in perception in general, and for that matter, ambiguous perception in particular [76]. The experimental research presented in this report addresses a case of variability between musicians and non musicians in the perception of an ambiguous pitch shift, thus raising the question of long-term modifications of the auditory system and its impact on ambiguous auditory signal.

Conclusion

In conclusion of this chapter, it appears that perceptual ambiguity is a ubiquitous phenomenon in general perception [178], that occurs in the perception of stimuli in different modalities [295, 61, 414, 334]. Despite the various paradigms used to study perceptual ambiguity empirically, several characteristics are recurrent, such as the perception of alternative interpretations for the same sensory signal [221, 334], the temporal dynamics of perceptual switches for continuous presentations [209, 295, 416, 202] and the contribution of sensory processes modulated by top-down influences [379, 165]. The exclusivity of competing interpretations is consistent with the idea that resolving ambiguity is a frequent perceptual task that hence requires fast and efficient processing [371]. However, the exclusivity of competing percepts have been recently challenged by findings of the build up of mixed percepts over the course of presentations [251, 188, 86], a phenomenon that could

reflect adaptation to a novel stimulus [188, 289]. Furthermore, exclusivity being grounded in ecological reasons, it is also possible that inter-individual variabilities in exclusivity of percepts reflect distinct strategies and needs for different individuals.

Chapter 2

Context effects and perceptual ambiguity

Objects, whether visual or auditory, are never perceived in isolation but instead, are integrated into a general context. This context may be simultaneous, in which case it corresponds to co-occurring informations, or it can occur prior to stimulation but whatever its synchronicity . Contextual influence is in itself a vast topic of research and its study encompasses a wide range of stimuli. Scientist have taken advantage of the methodological possibilities of ambiguous figures and implemented it in context effects paradigms [136, 158, 311, 165]. Ambiguous stimuli are very well-fitted tools to grasp the importance of contextual information, for their undetermined nature makes them very sensitive to prior input [190].

2.1 Simultaneous context

The sensory input that is being processed at a given moment is surrounded information, simultaneous or past. As we have in the previous chapter, the results from Intaite et *al.* suggest that contextual cues have a very divergent

impact of the processing of information, if they are prior, or simultaneous [165]. In this section, we will briefly review the main findings of simultaneous contexts. The effects of prior context will be more thoroughly discussed later in this chapter, as they are closely linked to the topic of this Ph.D.

When ambiguous visual objects are perceived simultaneously with unambiguous other objects, perceptual systems takes into account the unambiguous information to resolve the ambiguous one. Many instances of spatial context effects have been reported in the literature, in binocular rivalry [355, 137] and more complex objects, such as moving cylinders [122, 189].

Another source of simultaneous context arise from other modalities. Cross-modal interactions are known to help the processing of noisy [110] and ambiguous stimuli [334]. In a verbal transformation study, Sato et al. found that visual input of lips wording the different interpretations of the repeated phonetic sequence had an influence on the occurrence of perceptual switches: they were triggered so to maintain congruency with the visual input [322]. In a binocular rivalry study, Conrad et al. found that the duration of perceptual dominance of a motion stimulus was increased when it co-occurred with a congruent sound motion, and decreased otherwise [70]. Emphasizing the effect of a less studied sensory modality, Zhou & Chen demonstrated that in visual rivalry, olfactory input could influence the dominance duration of the congruent visual image [415].

Sensory information coming from surrounding parts of the scene [355, 137, 122, 189], or other modalities [322, 415], are therefore used to resolve noisy and inconclusive sensory input, maximizing the overall coherence of the scene.

2.2 Contrastive vs. assimilative

Context effects correspond to the influence of surrounding or preceding information on a given test stimulus. Some are termed "contrastive", for the perception of the subsequent stimulus is "shifted away" from the preceding

context. For instance, the prolonged presentation of a motion stimulus in one direction creates the illusory perception of motion in the opposite direction for a subsequent stationary stimulus [11]. The effect of context can also be "assimilative", in the sense that contextual information attracts subsequent perception to maximize congruency and similarity. This occurs for instance when a brief motion biases a subsequent ambiguous motion towards the same direction [172].

2.2.1 Distinct functional benefits

Although they have been traditionally studied separately in distinct studies, more and more studies have shown that a same stimulus can be biased in a contrastive or assimilative manner, with only slight changes in the context presentation methods [172, 165]. In Intaite et al., the perception of a Necker cube was contrasted from the prior prolonged exposition and attracted to the simultaneous presentation of the same unambiguous stimulus [165]. In a motion perception study, slightly manipulating the temporal parameters of presentation of a motion stimulus induced contrastive and attractive effect on the perception of a subsequent ambiguous motion [172].

The co-existence of two opposite effects on the perception of a same stimulus may subserve distinct functional benefits [336]. It was argued that contrastive effects result from enhanced responsiveness to novel information. By decreasing responsiveness to repeated stimuli, emphasized response to novelty could allow perceptual systems to better detect novel, hence critical stimulation [397, 200]. Therefore, contrastive effect could be the perceptual manifestation of the optimization of information transfer [116].

Assimilative effects would be the perceptual manifestation of the complementary phenomenon of perceptual stabilization. Indeed, objects properties tend to persist over time, although sensory input might vary due to external factors (such as masking [121]), making the recent past a good predictor of the present [99, 116]. Therefore, attractive effects might serve the purpose

to stabilize percepts, a coping process due to constantly changing low-level stimulus features [185, 336].

2.2.2 Auditory context effects

Contrastive effects

A contrastive effect in speech perception was identified over fifty years ago for the perception of formants [216, 225, 239]. In the speech signal, formants are defined as the peaks of spectral energy in the frequency spectrum of speech sounds. Vowels are mainly characterized by the two first formants, "F1" and "F2" and are ordered on a scale from low F1 and F2, vowel [u], to high F1 and F2, vowel [a] which play a crucial role in vowel recognition [112].

In an influential study, Ladefoged & Broadbent presented a test speech sequence composed of a b<vowel>t in which the two main formants of the vowel F1 and F2 were shifted so to have four test word conditions: A, B, C and D (top panel of Figure 2.2.1) [216]. The four test sequences were preceded by a context sentence whose formants were also shifted so to have six conditions of context, one control condition of non-shifted formants and five conditions of either formant one F1 or two F2 shifted up or down (bottom panel of Figure 2.2.1). The task consisted in identifying the test word as "but", "bit", "bet", "bat" or "but". The results showed that the test word "D" was perceived as "but" in 82% of cases and as "bat" in 18% of cases when preceded by context sequence one (no shift of formants). In contrast, word D was perceived as "but" in 38% of cases and as "bat" in 60% of cases when preceded by context sequence four (formant "F2" shifted down). The results hence demonstrated that the identification of vowel is influenced by the preceding context and that this effect is contrastive: when the preceding context possesses high-frequency acoustic energy, listeners are biased to categorize subsequent speech sounds as possessing lower-frequency energy.

TABLE II. The frequencies of the first two formants in the four test words.

Test word	Frequency in cps	
	Formant one	Formant two
<i>A</i>	375	1700
<i>B</i>	450	1700
<i>C</i>	575	1700
<i>D</i>	600	1300

TABLE I. Differences in the six versions of the introductory sentence: *Please say what this word is.*

Sentence version	Differences from sentence 1	Frequency range in cps	
		Formant 1	Formant 2
1	...	275-500	600-2500
2	F. 1. down	200-380	600-2500
3	F. 1. up	380-660	600-2500
4	F. 2. down	275-500	400-2100
5	F. 2. up	275-500	800-2900
6	F. 1. down F. 2. up	200-380	800-2900

Figure 2.2.1: Test sequence and context sequence in a study on contrastive effects in vowel identification. From [216].

This effect of spectral contrast in vowel identification was believed to reflect the strong implication of motor processes in vocalization [225] and hence, to be specific to speech perception [120].

However, further studies have reported cases of spectral contrasts in more generic processes of auditory perception. A series of studies have gathered evidences that even non-linguistic context influence the perception of subsequent speech context in a contrastive way [161, 232, 158]. By presenting a context sequence, here termed "acoustical history", made of pure-tones with frequencies in either a high or low range before a nine test speech sequences gradually shifting from /ga/ (low F2 and F3 formants) to /da/ (high F2 and F3 formants), they assess the influence of non-speech information on speech identification (left panel of Figure 2.2.2).

The results (right panel of Figure 2.2.2) indicate that the perception of the test stimulus as "ga" (low formants) or "da" (high formants) depends on

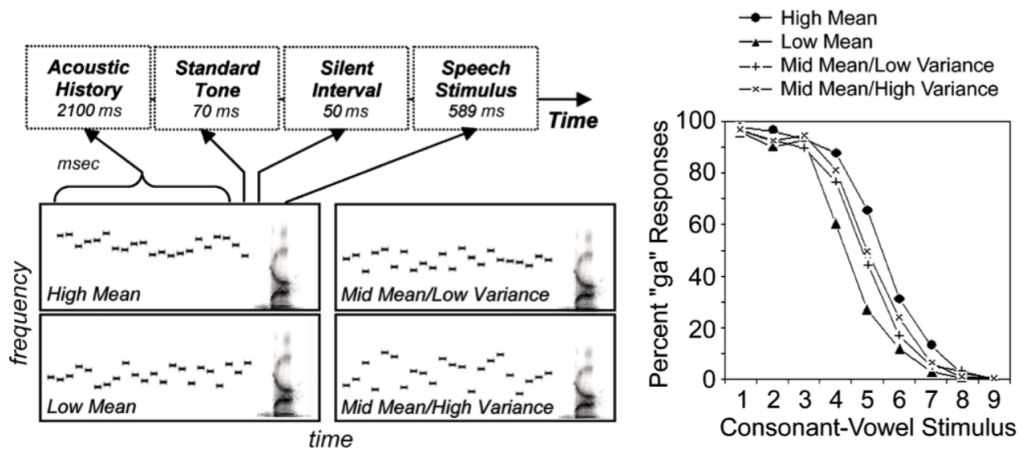


Figure 2.2.2: Left panel: schematic representation of experimental procedure: acoustic history with distinct frequency range is followed by a speech stimulus after 50ms of silence. Right panel: mean percentage of "ga" response in each test stimulus condition (consonant-vowel condition), for each context condition (different curves). From [158].

the preceding context. For instance for stimulus condition 4, listeners report hearing "ga" in 30% of cases when preceded by the low-range frequency context, and in 70% of cases when preceded by the high-range frequency context. Hence, when the acoustic history possesses a relatively high frequency range, the stimulus test is perceived as possessing relatively low frequency range.

Similar contrastive effects have been observed when a speech context precedes a non-speech test [359], and was even observed for non-speech sounds [362]. Stilp et al. presented context sequences spectrally manipulated so to be perceived as "french horn" or "tenor saxophone", followed by non-speech targets [362]. The results (Figure XX) echoes findings on enhanced spectral contrast from Holt et al.: listeners were more likely to report "saxophone" for the target stimulus when preceded by context spectrally manipulated to imitate French horn, and vice versa. Altogether, these findings suggest that both speech-specific and auditory-generic processes are involved in contrastive speech context effects [160, 347].

In sound localization, a localization "aftereffect" has been reported: the

subjective localization of a sound is shifted away from the localization of a preceding localized sound [75, 175]. In Dahmen *et al.*, the statistics of the acoustical context was varied in three conditions: the mean Interaural level differences (ILD, a localization cue for binaural listening [141]) of broadband noise sequences was fixed at -15dB, 0dB or 15dB [75].

Listeners indicated whether they perceived the position of a static stimulus with varied IDL presented immediately after the context to be on the left or right. The context effect was characterized in terms of psychometric functions plotting the percentage of “left” responses as a function of ILD conditions for the different context conditions (Figure 2.2.3). The results show that changing the mean of the ILD in the context shifted the perceived localization of a stimulus away from the mean.

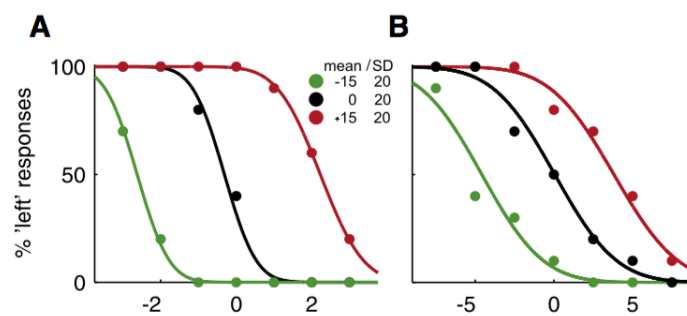


Figure 2.2.3: Individual results of two listeners. The proportion of “left” response is plotted against the ILD condition of the test tone for the different context conditions. From [75].

In an influential study, Shu *et al.* reported what they termed a spectral motion “aftereffect” [346, 176, 126, 240], in reference to the motion aftereffect, documented in vision and whereby the prolonged exposition to a moving stimulus induce a shift of motion perception in a subsequent stimulus [11]. In the spectral motion aftereffect (also termed “spectral-motion contrast”), the presentation of a narrowband noise with moving center frequency either upward or downward along a frequency axis (or “frequency glides”) is presented for 2-3 minutes to listeners. A subsequent brief noise was presented immediately after the adapting stimulus and listeners were asked to report

wether the test stimulus was going "up" or "down" in frequency. The subjective spectral motion was affected by the context in a contrastive manner: when the context frequency motion was going upward, a upward test stimulus was perceived as stationary and a stationary test stimulus was perceived as going downward, and reciprocally for downward motions.

These findings were supported by further studies that observed similar contrastive effects for even shorter presentation durations [6, 392]. In Wang & Oxenham, the spectral motion aftereffect was observed with for non-speech and speech target stimuli, with short context sequences (100 and 500ms). The effect was contrastive and frequency-specific: the perceived direction of the test stimulus (a falling, raising or flat tone of 50ms) was shifted away from the context direction frequency shift and this effect was stronger when the test and context stimuli had the same frequency range, consistently with Alais et al. [6] (the effect of context is stronger in middle panel of Figure 2.2.4 than in other panels).

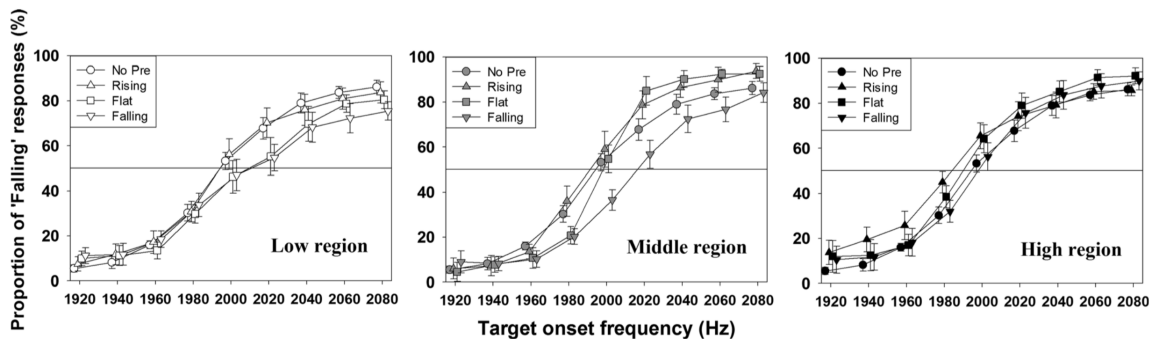


Figure 2.2.4: Averaged proportion of "falling" reported direction plotted as a function of the onset frequency of the test tone, for each context condition (different curves) and each frequency range of context (different panels). From [392].

Assimilative effects

In contrast, only few instances of assimilative effects have been reported in the auditory modality. In the verbal transformation effect, introduced in

the previous chapter, a coupling of perceived phonetic sequences has been observed [98]. This constitutes a case of perceptual hysteresis, one of the hallmarks of attractive effects [210].

In the Shepard tone paradigm, the perception of an ambiguous tritone interval of two Shepard tones was also found to exhibit hysteresis [63, 132]. As emphasized by Hock et al., hysteresis effects may be difficult to interpret, for hysteresis in response patterns may be result from perceptual of response biases.

In an elegant study, Chambers & Pressnitzer circumvented this methodological issue and reported hysteresis in the pitch perception of Shepard tones [63]. Listeners had to listen to an interval of two tones, a standard (black) and a comparison (grey), and report which of the tones was higher in pitch. Intervals of standard and comparison tones were presented in sequences of either increasing or decreasing order. But in each interval, the presentation order between standard and comparison was randomized (see left panel of Figure 2.2.5). This way, the direction responses ("up" or "down") were independent of the order condition. From the response direction ("up" or "down") and the standard position in the interval (first or second), a proportion of standard perceived as higher $P(\text{SH})$ was computed for each interval size.

Listeners had to listen to an interval of two tones, a standard (black) and a comparison (grey), and report which of the tones was higher in pitch. Intervals were presented in sequences of either increasing or decreasing order (increasing: S-C interval from 1 semi-tone step to 11 semi-tone steps, decreasing: S-C intervals from 11 semi-tones steps to 1 semi-tone step). But critically, in each interval, the presentation order between standard and comparison was randomized (see left panel of Figure 2.2.5). This way, the direction responses pattern ("up" or "down") was independent of the order condition. From the response direction ("up" or "down") and the standard position in the interval (first or second), a proportion of standard perceived as higher $P(\text{SH})$ was computed for each interval size.

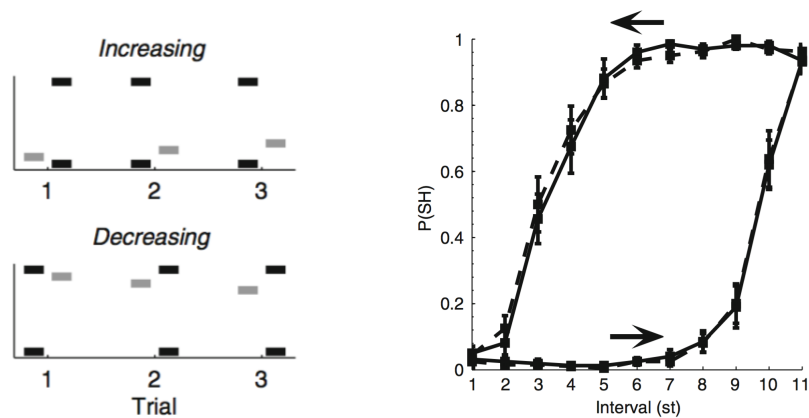


Figure 2.2.5: Left panel: schematic representation of increasing and decreasing order of intervals between standard (black) and comparison (grey) Shepard tones (y-axis in frequency log). Right panel: averaged proportion of standard tone perceived as higher than comparison tone as a function of the interval size for the increasing and decreasing order conditions. From [63].

A hysteresis effect was observed, whereby the perceived height of the standard tone was influenced by the preceding context: remarkably, the standard tone perceived height in comparison to the comparison tone drastically differed between the two order conditions (see right panel of Figure XX). However, this hysteresis effect could not reflect decision hysteresis, since the randomized presentation prevented that increasing order condition resulted in all "up" responses and vice versa for the decreasing order condition. Therefore, this study reported a strong hysteresis effect in the perception of pitch height that was perceptual in nature, since it could not be attributed to hysteresis in response patterns.

Another attractive effect was observed in a study investigating the effect of prior trials on a two-tone discrimination task [304]. Listeners had to report which of two tones was higher in a sequence of trials of intervals whose sizes was adapted with a stair-case procedure. Consistently with the Bayesian framework, expectations derived from previous trials influenced the responses: small magnitudes relative to the distribution of stimuli used in the experiment were overestimated while large magnitudes relative to this

same distribution were underestimated, a phenomenon termed "contraction bias" [13].

Finally, in auditory scene analysis, Snyder et al. reported an assimilative context effect in the perceptual organization of "ABA" bistable streams [352]. For certain temporal conditions, they observed that listeners tended to report "2 streams" in higher proportions when they last reported "2 streams" at the end of the context period.

2.3 Temporal dynamics of context effects

Prior information has a very influential effect on the perception of subsequent noisy or ambiguous stimuli [190]. Temporal parameters of context effects seem to be very important to trigger context effects. Durations of presentation may be critical to trigger contrastive or assimilative effects and intervals of "blank" between context and test stimuli have different effects on the decay of context bias. These elements are highly informative in the underlying neural mechanisms. In the next section, the temporal dynamics of different context effects will be reviewed and discussed.

2.3.1 Long context durations

Long presentation duration of context effects (on the order of seconds) have been reported to typically result in a contrastive effect. This has been applied and successfully replicated in many visual paradigms [45, 172]. In binocular rivalry, the long presentation of an image makes it more likely to be suppressed when the second image is added [45]. Similarly, in visual bistability, the perceived direction of an ambiguous motion is shifted away from a long previewed context motion [11, 172].

Auditory contrastive effects, such as those reviewed in the previous section were also triggered by long presentation durations of context. In the contrast

effect of vowel identification, studies report presentation durations on the order of seconds [216, 161, 232, 158]. In Ladefoged & Broadbent and Stilp *et al.*, the context sequence is a sentence of approximately 1 second duration [216, 362]. In Holt *et al.*, the context sequence used to bias the speech stimulus is a sequence of tones of 2,1 s duration [158]. The sound localization contrast effect has been reported to follow a context sequence of 1s [75].

Similar temporal dynamics are observed in contrastive effects in other sensory modalities. In a study investigating rivalry in the olfactory modality, Zhou & Chen found that long exposition to a unique fragrance induced a contrastive effect during a further exposition of competing fragrances in the two nostrils [414] and a similar contrastive effects was also found in a study exploring the ambiguous motion of a tactile stimulus [61].

All these instances may be indicative of the same phenomenon: adaptation resulting from long durations of presentation. Adaptation is a phenomenon believed to result from fatigue-like mechanisms of populations of neurons, by which the responsiveness of neurons is decreased for subsequent presentations [3]. It has been documented in vision [200], audition [56, 49, 382], and olfaction [290]. Contrastive context effects may therefore result from the decrease of responsiveness in certain population of neurons, due to adaptation to long durations of presentation. This view is consistent with findings suggesting that the strength of the contextual information is also critical in inducing contrastive context effects [45].

2.3.2 Short context durations

In contrast, brief context durations (in the order of milliseconds) have been associated with attractive context effects [286, 306, 172, 45]. In visual motion, Pinkus & Pantle tested the effect of context presentation duration on the perception of an apparent motion resulting from the sequential presentation of three static frames of gratings [286]. The transition of phase between frames determined an apparent motion, a transition phase of 180

degree resulting in an ambiguous motion. The motion step between frame 1 and frame 2 had an effect on the perceived direction of the ambiguous phase step between frame 2 and frame 3. The amount of visual motion priming (the proportion of trials in which the ambiguous motion was perceived as going in the same direction as the previous unambiguous motion) was plotted as a function of the context duration, here the presentation of frame 2, for two contrast conditions and three spatial-frequency conditions of gratings (Figure 2.3.1).

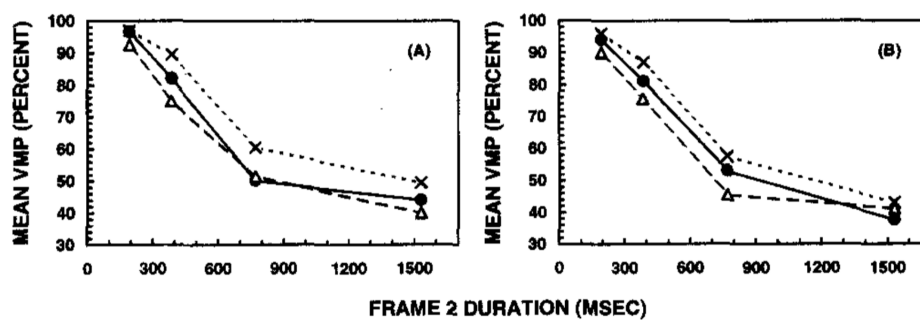


Figure 2.3.1: Proportion of visual motion priming as a function of the duration of context presentation. Each line corresponds to a different spatial-frequency condition of grating and each plot corresponds to a different contrast condition. From [286].

The results demonstrate that visual motion priming is established for very short context presentations (100ms) and rapidly decays as the context duration increases. This effect contrasts with the motion aftereffect which correspond to the opposite effect of the context and which is triggered by longer context presentation durations [11]. This perceptual priming could result from the increased responsiveness of neurons observed after very brief stimulation [152].

However, recent findings in the auditory modality suggest that short durations also trigger contrastive effects. In an early study, a spectral-motion contrast, or spectral motion aftereffect was reported to follow 2-3 minutes of context presentation [346], consistently with adaptation dynamics observed on auditory neurons [235, 69].

But recently, two studies reported the same effect with much shorter context durations (100ms-500ms) [6, 392]. Using an inter-trial approach, in which the effect of prior context is measured continuously, so that each trial is both the test of the preceding trial and the context of the subsequent one (see method in Figure 2.3.2), Alais et al. observed a contrastive effect on the perceived frequency motion established by very brief context sequences [6].

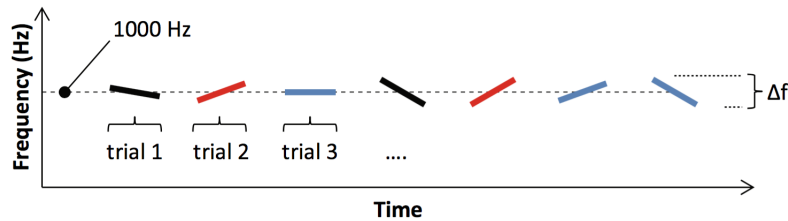


Figure 2.3.2: Inter-trial procedure in a spectral motion study. Listeners are asked on each trial to report the direction of the glide. A current trial (t) is both the context of the subsequent trial ($t+1$) and the test of the preceding trial ($t-1$). Colors code the context condition for each trial: red, downward context, black, flat context, blue, upward context. From [6].

The perceived direction of tone glides going upward or downward was collected on each trial. The results revealed that the subjective motion of a trial was influenced by the two preceding trials ($t-1$ and $t-2$). That is, each tone glide of 100ms had a contrastive effect on the subjective perception of the subsequent two tones.

In a study by Wang & Oxenham, a similar spectral-motion contrast was observed for both non-speech and speech test stimuli [392]. Interestingly, a context sequence of 500ms failed to elicit the context effect: only test stimuli preceded by a 100ms context stimulus were biased.

Emphasizing the difference between the spectral-motion contrast effect observed in their study from the spectral contrast effect reported elsewhere [216, 161, 232, 158], these findings suggest that those unusually short temporal dynamics could reflect the specific adaptation time-course of neurons tuned to frequency shifts [84, 83].

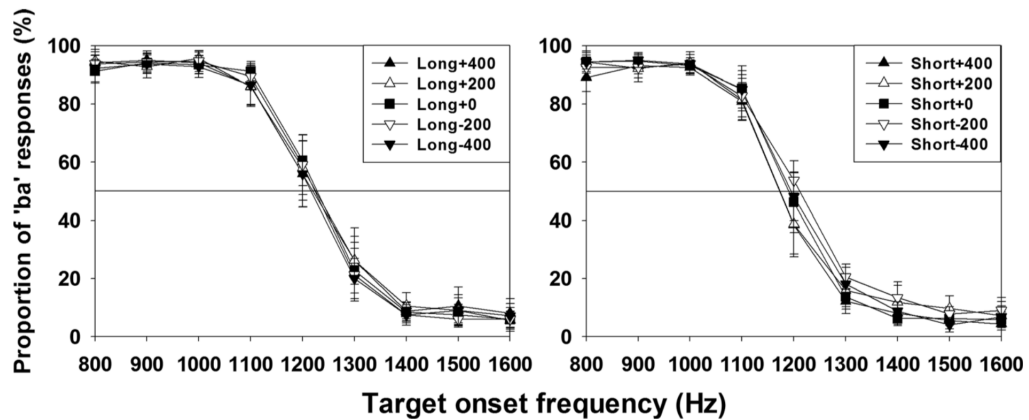


Figure 2.3.3: Proportion of "ba" responses following a context sequence of various durations and as a function of the frequency of the target stimulus. From [392].

2.3.3 Gap duration

Gap duration refers to the time of "blank" or "silence" following the context and preceding the test [158, 172]. In intermittent presentation of ambiguous stimuli, the gap that separates consecutive presentations have been reported to lead to assimilative effects: ambiguous stimuli tend to evoke the same percept on many consecutive presentations, a phenomenon referred as perceptual stabilization (see Figure 2.3.4) [222, 236, 234].

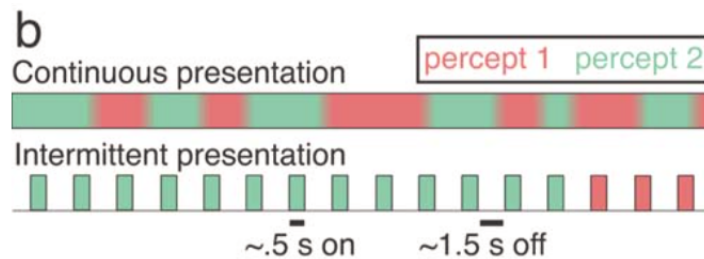


Figure 2.3.4: Perceptual alternations in continuous presentation conditions and perceptual stabilization in intermittent presentation conditions. From [44].

This assimilative effect may be triggered by facilitatory effect of the past perceptual dominance on subsequent perception that slowly grow stronger

over multiple presentations [277, 44]. Interestingly, the time-course of this assimilative effect is extremely slow: the effect has been observed to persist for a minute [234] to even dozens of minutes [46].

In binocular rivalry, Brascamp et al. introduced a blank of varied duration between stimulation of one of the two stimulus and simultaneous presentation of two stimuli in different eyes [45]. For short presentations, an assimilative effect was observed: the dominant percept during the context phase tended to be maintained, as discussed in the previous section. They also observed an effect of the duration of the blank: short durations (0 and 12ms) tended to weaken the facilitative effect observed for longer blank durations (> 50ms). In an ambiguous motion study, Kanai & Vestraten observed a similar effect: the assimilative effect caused by an ambiguous motion context was increased by the duration of the blank [172]. In the auditory modality, the effect of preceding context was also found to persist over long gap durations, in speech categorization [158] and auditory scene analysis [352].

As pointed out by Maier et al. such slow temporal decays are informative of the nature of the neural processes involved. Indeed, this persistence suggest that the locus of the underlying neural processes is not likely to be found in the primary sensory areas [234]. This hypothesis was supported by results observed in ambiguous motion perception. When an ambiguous motion was used as prior exposition, the assimilative effect observed decayed very progressively. As they point it out, the fact that the slow decay was observed only for ambiguous stimuli suggests that the effect is most likely linked to percept-based biases, occurring beyond the primary sensory areas [172].

Snyder et al. reported a similar effect in the auditory modality [353]. They investigated the effect of context "ABA" sequences with different degrees of ambiguity on subsequent ambiguous "ABA" test sequences. When the context sequence was not ambiguous, a contrastive effect occurred as listeners tended to report the alternative mechanism, an effect seemingly due to suppressive mechanisms [353]. However, when the context sequence was itself ambiguous, an assimilative bias occurred (see Figure 2.3.5). In this fig-

ure, trials are sorted according to whether they were reported as two streams (dashed line) or one stream (solid line) at the end of the context sequence. An assimilative bias corresponded to the fact the listeners who heard two streams at the end of the context sequence tended to maintain this percept during the test sequence and *vice versa*.

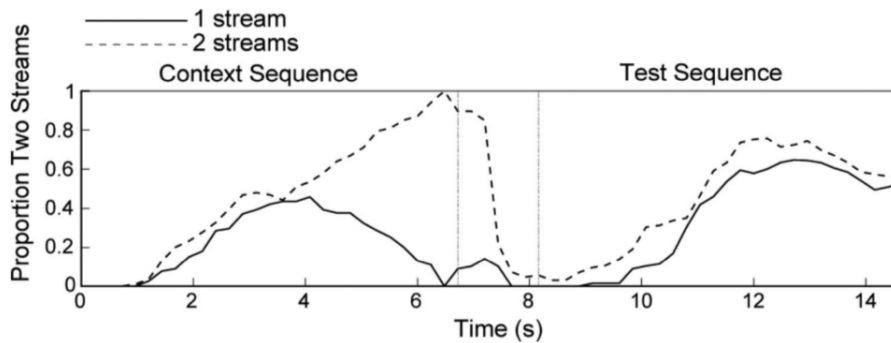


Figure 2.3.5: Proportion of time trials heard as streaming for the context period (left portion of panel) and test period (right portion of panel). Trials are sorted according to whether they ended on a two-stream percept (dashed line) or a one-stream percept (solid line). From [352].

2.3.4 Systematic examination of temporal dynamics

A brief review of the literature of context effects revealed that temporal dynamics play a crucial role in the influence of contextual information. To over-simplify, long presentation durations are associated with contrastive effects and short presentation durations with attractive effects. The duration of the blank or silence between context and test also seems to play a critical role [172, 268, 46].

However, the disparity of methods, stimuli and procedure makes it difficult to interpret the disparate findings. In order to observe directly the transition between contrastive and assimilative effects, some studies have varied systematically the temporal dynamics of context effects.

Kanai & Verstraten conducted series of experiments in which they sought to systematically explore the effect of temporal parameters of context presentation on the perception of an ambiguous visual motion [172]. The ambiguous probe stimulus was an alternation of luminance gratings phase-shifted by 180 degrees that biased participants to perceive the motion either toward the right or the left. It was preceded by an adapting stimulus which was an alternation of gratings whose phase-shift strongly favored one or the other direction. A schematic representation of the procedure is displayed in the left panel of Figure 2.3.6. The duration of the adapting non-ambiguous stimulus and the time separating it from the ambiguous probe were systematically varied. Results are displayed in the right panel of Figure 2.3.6. They show that a brief context followed by a short gap result in an assimilative bias, whereby the motion perceived in the probe stimulus is congruent with the adapting one. For longer context durations, the effect of the adapting stimulus was contrastive. Both assimilative and contrastive bias tended to decay for long gap durations.

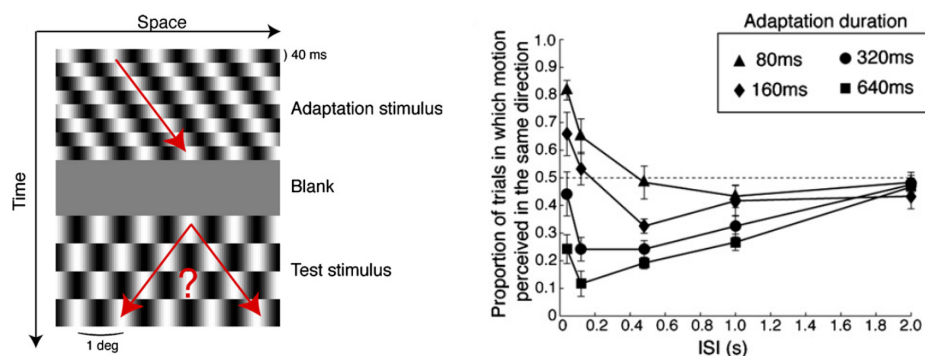


Figure 2.3.6: Left panel displays the stimuli and procedure in Kanai et al. Right panel displays the results. The proportion of motion congruency between context and test stimuli are displayed as a function of different ISI (on the x-axis) and for different context durations (different lines). From [172].

Similar findings were found in binocular rivalry. Brascamp et al. conducted an experiment in which duration and contrast of an adapting stimulus were varied systematically [45]. They found that the type of effect induced by a

context depended on temporal parameters. Brief exposure of one of the two stimuli enhances its subsequent dominance (facilitation). Longer exposure durations (around .5 s) had the opposite effect (suppression) (see Figure 2.3.7). This was particularly pronounced for the highest contrast level.

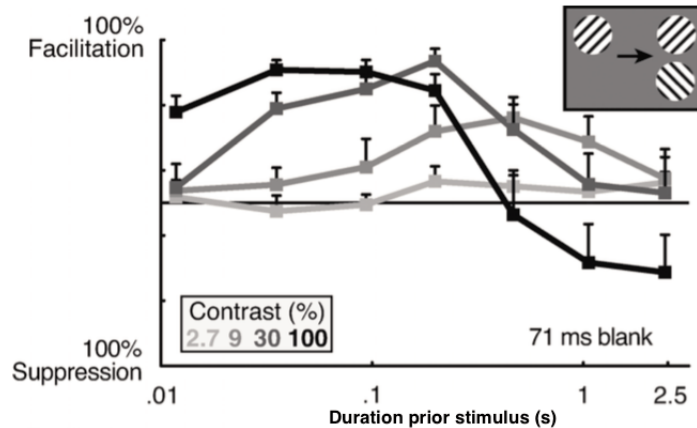


Figure 2.3.7: Amount of contrastive (suppression) or assimilative (facilitation) effect as a function of context presentation duration for different contrast levels. Brief exposure enhances its subsequent dominance (facilitation), but exposure durations over about .5 s have the opposite effect (suppression). From [45].

Altogether, these findings demonstrate that prior exposure may result in different context effects depending on temporal parameters used for context duration. Several interpretations can account for the existence of both types of effects on the same stimuli. Some have explained attractive and contrastive effects with a single mechanism [128], some with two distinct phenomena located in the same early stages of processing [268], or distinct stages of processing [336].

Noest et al. proposes a model in which pre-exposure induces a decrease of sensitivity of responsiveness, together with a small facilitation, occurring at a single stage of low-level processing [268]. While the decrease of responsiveness tends to build up slowly over time, the enhanced sensitivity decays very rapidly. The temporal dynamics are therefore crucial to determine the type of influence on subsequent stimuli. Duration of context presentation

will be critical to determine which effect will have prior stimulation: long presentations would allow adaptation to build up and short presentation will allow priming [172].

Conclusion

In the light of the studies discussed in this chapter, it appears that contextual information, whether preceding or simultaneous [165], plays a tremendous part in what determines how a particular stimulus is perceived. Ambiguous stimuli, being inconclusive by nature, constitute great tools to observe the specific processes of context information integration [190]. Temporal parameters of context presentation and inter-stimuli blank can lead to very different perceptual outcomes [172, 45]. Therefore, systematically exploring the time-course of context effects may gather very informative elements on the type of neural processes at play.

Chapter 3

To be or not to be a musician

Music is a phenomenon observed in every known human culture [247]. However, the degree with which individuals in a society engage in musical activities greatly varies across cultures. Some cultures lack of the notion of a hierarchy in musical abilities among individuals whereas others consider musicianship as a specific competence [36]. In Western cultures, musicians are typically considered holders of a special faculty which distinguish them from non-musicians. They are trained from an early age in specialized schools to learn how to play an instrument and are also taught various theoretical matters such as harmony, history of music and analysis. This intense practice (several hours per day during many years), together with the fact that it most often begins early in life, makes musicians a very interesting population: their brains constitute a fine experimental object for the investigation of how long-term experience changes brain structures and cognition [146, 409]. This chapter will first face the task of defining music as a specific activity and musicians as a specific population. Once we have defined who they are, and what they are trained for, we will address the question: are musicians different from non-musicians from a cognitive standpoint?

What is music and what do musicians learn ?

Finding a definition that would describe music with a satisfactory degree of specification and yet, that would embrace all its stunning variability amongst and across cultures, appears to be challenging. Hence, for the purpose of our analysis, we will focus on Western tonal music and settle on the minimalist definition suggested by Rodriguez, according to which "Music consists of sound organized in time, intended for, or perceived as, aesthetic experience (Rodriguez, 1995)" cited in [100]. Those sounds organized in time vary along the main acoustical dimensions (i.e. pitch, loudness and timbre) and constitute the first level of description and analysis that is referred to as musical *surface*. On the other hand, musical *structure* corresponds to the fact that distinct musical events fulfill different structural functions beyond the immediate sounding qualities. Those two distinct levels of organization may require different levels of cognitive processing, as we might expect the level of musical *surface* to rely on mostly perceptual processes and the level of musical *structure* to recruit higher-level cognitive processes [223].

As a musician, what abilities do one acquires through training that make him or her able to interact with music as a professional? A first approach is to examine what education is provided by institution devoted to train musicians, in order to understand what specific skills they acquire. First, a musician most often must know how to play an instrument, notwithstanding the fact that his intention is to become a musicologist, a conductor, a composer or any other musical profession that does not involve performing [349]. Playing a musical instrument requires a certain dexterity, specific to each instrument, that can be only achieved through years of daily practice, estimated in total at 10,000 hours [206]. Moreover, as a large part of Western music is written for small and large ensembles of instruments -ranging from the sparing sonata for which only two instrumentalists are necessary, to the full symphonic orchestra which sometimes require as much as one hundred musicians-. Being able to play in tune and rhythm with others appears to be a critical ability. In addition to that, a much more theoretical training is also completed, for the

musician has to learn the basics of music theory and harmony, also has to be able to analyze a music score -very much like one would learn to comment a poem- and to describe a musical excerpts just by listening to it, all those tasks being challenging on an intellectual level [40].

Assessing musical abilities

Musical achievement is not only modulated by the intensity and precocity of musical training [40]. As in any other life-time activity, such as sports and mathematics, musical achievement also depends on self-efficacy [41, 57], self-regulation [417], and time spent on deliberate practice [109]. Furthermore, it is very difficult to precisely describe the amount of musical education a musician received, for there are very different strategies of music educations that may impact certain musical abilities [340, 182]. Finally, musicianship is a multi-faceted ability that includes, aside from playing an instrument, musical understanding, appreciation and evaluation [147], which makes the very definition of "musical expertise" and its measurement, even more challenging [224]. Here, we will briefly review the main tools of measurements that have been used in scientific studies.

Since the first scale for measuring "musical talent" was imagined by Seashore in 1919 [337], numerous scales have been conceived. Most of them rest on a same-different paradigm in which listeners have to detect subtle differences occurring along different dimensions of musical surface or structure (for a recent review, see [262]). Yet, because they engage musical material and procedures very similar to what musical institutions use to teach, those tests favor musicians that have received a formal training, thus failing to fairly assess the full scope of musical abilities. It seems also that some crucial musical abilities are not directly tested through those evaluation. For instance, the stimuli used are never composed of several instruments playing together and listeners are never asked to analyze the timbre of a chord (e.i. which instruments are simultaneously playing together), although this ability

may be critical in some musical activities, such as music chamber practice or analysis of orchestral music.

Law & Zentner legitimate this focus on simple musical stimuli by arguing that the processing of structural musical "atoms" is a pre-condition for the successful decoding of more complex musical content and hence, the ability to process musical "atoms" would be a valid proxy for assessing higher-order musical skills [217]. Aside from assessing musical abilities through explicit testing, questionnaires were also developed, with the aim to assess the multidimensional and continuous nature of musicianship, which extends the concept of "musicianship" beyond formal music training and performance factors, and captures both quantity and quality of music production and reception [65]. Current research seeks to develop this fine-grained analysis of the multi-faceted nature of musical expertise through explicit testing [262].

Despite of the aforementioned attempts to robustly and precisely quantify musical expertise, there is a lack of homogeneity in criteria applied to musicianship across studies, with years of training ranging from at least two years to at least 6 years [?]. The explanation may be that the method to best define musical expertise is still a matter of debate [224] and can yield to a large scope of more or less sophisticated approaches. But ultimately, years of musical training constitute an arguably coarse but satisfying enough measure of musical expertise as it was proven to be highly correlated with musical expertise indexes across various scales [390].

3.1 Makes (almost) no difference

The origins of musical expertise has been largely discussed and remain a matter of debate [242]. If, as it has been postulated, music proficiency is rooted in innate predisposition [381] that have been selected through evolution for its adaptive social benefits [73], a high degree of musical competence should be present in the general population. As a matter of fact, non-musicians

are also exposed to music on a daily basis. This regular musical stimulation, together with the remarkable propensity of human beings to internalize regularities from the auditory environment [320] should interact with the innate predisposition [381] and result in a high degree of musical sophistication in all individuals, regardless of their musical training [33, 294, 293, 32].

3.1.1 Musical tension

In line with this, several studies designed to investigate the effect of musical expertise on the processing of Western music have suggested that non-musicians exhibit results that are surprisingly very similar to those of musicians (for a review, see [33]). In a series of studies investigating musical expectancies throughout a melodic progression, the results showed no difference between musicians and non-musicians [208]. The degree of consistency between the two groups suggest that musical expectancies are built largely independently from formal musical training. Concordant results arose from a series of studies conducted by Bigand *et al.*. These studies have taken advantage of the distinction between musical *surface* and musical *structure*, showing that non-musicians could correctly identify a musical structure despite the variations of musical surface elements [26]. Although musicians performed significantly better than non-musicians in this task (e.g. 72% versus 58% of correct response), the fact that non-musicians performed above chance demonstrate that they represent musical information on an abstract format, despite their lack of formal training. In another study, Bigand *et al.* sought to evaluate whether the effect of elements that determine tonal hierarchies (e.g. how pitch and rhythm influence stability and closure in a musical excerpts) are modulated by musical training [27]. To this aim, they presented listeners with two melodies, T1R1 and T2R1, very similar regarding notes and rhythms (see top panel in Figure 3.1.1). Yet, the tonal functions are radically different from one melody to the other, for one is written in A minor while the other is written in G major. Hence, individual notes carry a very different degree of stability, defined by their relation to

the tonic and thus different in the two melodies. For instance, note 8 is a tonic T2R1, thus carrying great stability and a subtonic in T1R1, thus carrying a great tension. Musicians and non-musicians were asked to report the "musical stability" perceived on each note (which is linked to the hierarchical proximity of the considered note from the tonic) on a 7-steps scale.

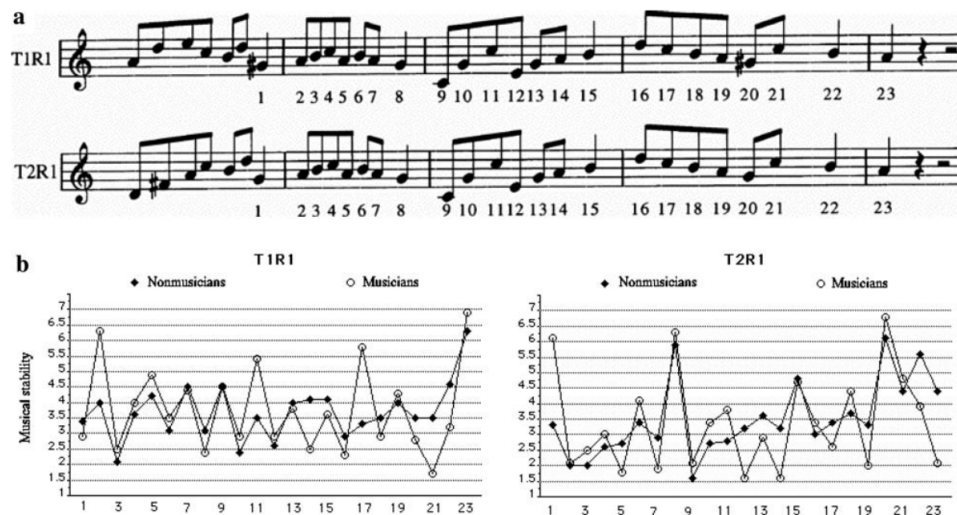


Figure 3.1.1: From [33].

The bottom panels in Figure 3.1.1 display the results for the two melodies and the two groups. The results suggest that both melodies received different stability rating patterns although they shared musical surface features thus suggesting that both musicians and non-musicians were sensitive to the key-change embedded in the two melodies. Moreover, ratings between the two groups were significantly correlated and on the whole, the pattern of data of both groups was very similar. The strong correlation between musicians and non-musicians have then been successfully replicated and extended to paradigms in which chord progression sequence were used instead of melodies [30, 29]. Altogether, those results suggest that tension-resolution dynamics in music are effective for both musicians and non-musicians and rely on abstract representations of musical structures elements.

3.1.2 Musical expectancies

The formation of musical expectancies might play a key role in music listening as it has been suggested that musical emotions arise from the composer's talent to frustrate or fulfill our building expectancies in order to create aesthetic sensations [164, 248]. On the other hand, it is conceivable that anticipation in music is developed through training. This would imply a strong advantage of musicians over non-musicians. Given the importance of harmonic structure in Western tonal music, expectancies derived from the harmonic context should constitute a critical aspect of music cognition [22, 23, 31, 377], and appear relatively early in life, regardless of musical education [325].

Mostly based on a priming paradigm [294, 23, 21, 35], several studies were conducted to investigate this point. A priming paradigm consist in presenting a probe or target stimulus -in this case, a tone or a chord- preceded by a context sequence that facilitate its processing to a varying degree (as illustrated in Figure 3.1.2). The task is to report a deviation in either timber, loudness or tuning on the target stimulus, as accurately and fast as possible. The critical point is to determine whether the preceding context facilitates the processing of the target, which will result in either higher performance or shorter response times. Bigand *et al.* presented a pair of two chords preceded by a very congruent context or a less congruent context. This way, the target stimulus would be very expected in the first case and less expected in the other case. In this experiment, the last chord was a consonant or a dissonant chord (by adding an extra tone to the triadic chord). Participants were asked to report whether the last chord was consonant or dissonant.

The results are displayed in Figure 3.1.3. They show that, although musician listeners usually performed better and faster, in the expected condition (e.i. condition for which the context is "congruent" with the target), context facilitate the processing of the target tone for both groups, as assessed by higher performance and shorter response times.

As pointed out by the author, the critical point in the results is that, despite

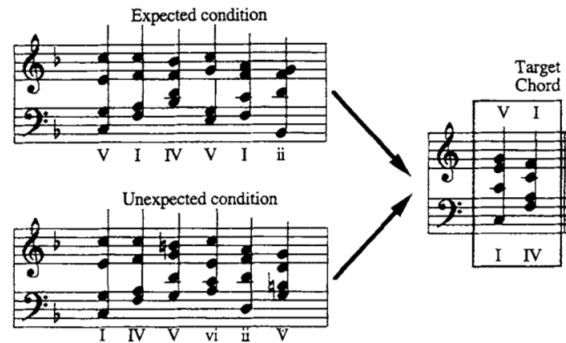


Figure 3.1.2: Example of two different sequences preceding two chords, one facilitate the processing of the target chords while the other does not. From [28].

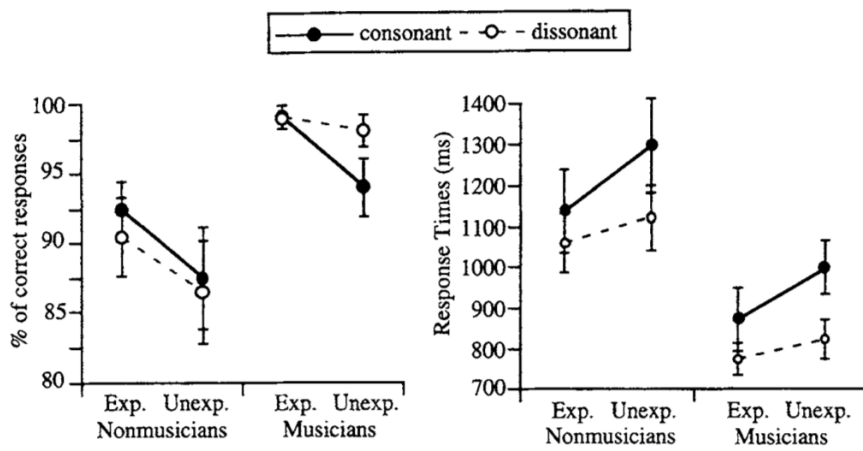


Figure 3.1.3: Percentage of correct response (left panel) and response time (right panel) averaged across sequences for the two target chord conditions (dissonant in dashed line and consonant in solid line) and for the two groups, musicians and non-musicians. From [28].

the group effect reflected in the two measures, the size of the priming effect revealed by the two conditions is similar for the two types of listeners. This suggests that non-musicians are also sensitive to the harmonic function of chords defined by the musical context. Those results were consistently replicated; in various following studies, using different paradigms, it was found that the size of the priming effect was equally pronounced in musicians and in non-musicians [34, 294].

In a further study, Bigand et al. sought to determine the level of processing of musical context priming [32]. The question arises because chords that typically end a syntactically correct chord sequence (i.e. tonic chords) contain *de facto* more tones in common with the contextual sequence than chords that are less stable at the end (i.e. fifth chords), thus favoring a sensory account of harmonic expectancies [330]. But on the other hand, harmonic priming could also result from the internalized knowledge of a listener's of the hierarchical relations between chords in a music piece [22, 23, 28], after implicit acquisition of the regularities in Western tonal music [374, 373]. To disentangle the sensory vs. cognitive roots of harmonic expectancies [68], Bigand et al. by manipulating the harmonic function of the target chord (cognitive factor) and the number of notes shared by the context sequence and the target chord (sensory factor).

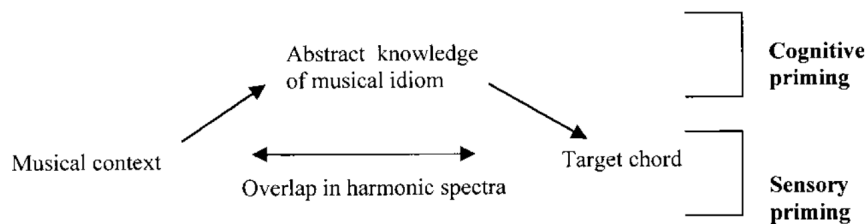


Figure 3.1.4: Sensory vs. cognitive account of harmonic priming. From [32].

The results showed that listeners were faster and more accurate in processing the target chord that had a higher position in the hierarchy (i.e. tonic chord), independently of the number of tones shared with the context and regardless of the group. Both types of listeners, musicians and non-musicians were

mostly influenced by cognitive rather than sensory priming.

This effect was then replicated with a different experimental setting [35] and on six-year-old children [325] and extended to electrophysiology studies [293]. Furthermore, a study exploring the integration of large-scale musical structures (such as tonality) in musicians and non-musicians revealed that musicianship does not provide a strong advantage: both groups failed to integrate local harmonic structures into global ones [375] (for a review, see [376]).

Electrophysiological recordings of neural response to syntactic musical irregularities opened a new avenue in the inquiry of music cognition and the effect of musicianship. The experimental protocol used is somewhat similar to those used in the aforementioned studies in that it rests on a priming paradigm ; the response to a probe tone, more or less congruent with a preceding context is measured in musicians and non-musicians. In response to a syntactic violation, an event-related potential (ERP) named the early right anterior negativity (ERAN) was identified (see Figure 3.1.5) and considered as the musical counterpart of the well-known Mismatch Negativity (MMN) (for a discussion on that matter, see [193]). The ERAN is supposed to reflect the processing of abstract and complex regularities carried by musical syntactical structures and results indicate that it builds up with age [167].

Koelsch *et al.* conducted several studies investigating the effect of musical expertise on the ERAN. They revealed that it was larger in professional musicians [197] and amateurs [196] than in non musicians (as illustrated in Figure 3.1.6). Yet, in the first study, the difference between the two groups was very small and barely reached the threshold of significance and in the second study, the group differences did not reach statistical significance. An fMRI study revealed a small but significant difference between adults and eleven-year-old children [194].

Other ERPs related to music processing have been observed [307, 293, 192, 358]. In Poulin-Charronat *et al.*, an ERP study was conducted to assess the contribution of sensory and cognitive priming in chord processing [293]

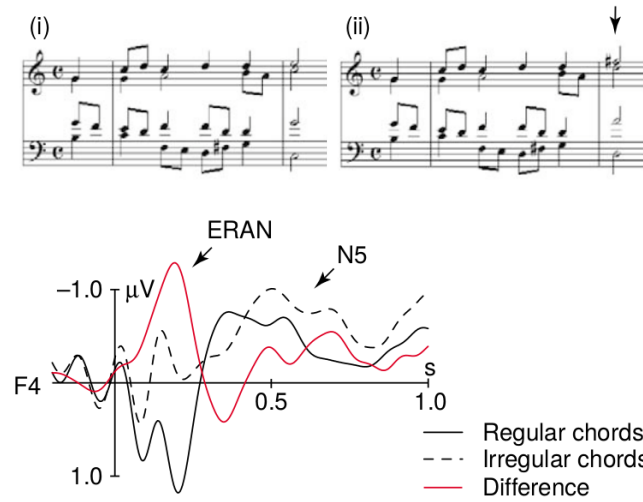


Figure 3.1.5: Top panel displays examples for syntactically correct chord progression in sequence (i) and incorrect one in sequence (ii). The bottom panel displays the ERPs elicited by the final chords of these two sequence types (recorded from a right-frontal electrode site from 12 subjects). From [192].

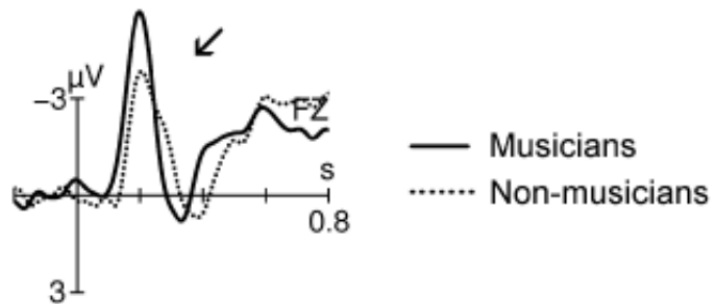


Figure 3.1.6: Difference in ERPs (expected chords subtracted from unexpected chords) elicited in musicians (solid line) and non-musicians (dotted line). From [193].

(similarly to [32]). The ERP was the N5, reported to reflect processes of harmonic integration [198, 195]. Consistently with previously observed behavioral findings, they demonstrated that the less stable chord target elicited a larger ERP than the more stable chord than the more stable chord, even though it was previously played in the context sequence. Besides, this effect was more pronounced for expert musicians than non-musicians, suggesting that the abstract representation of complex harmonic rules is still modulated by musical education.

On the whole, concordant evidence gathered from different behavioral and electrophysiological findings reveal that musicians exhibit a small advantage in processing complex musical structures when compared to non-musicians. However, this advantage should be considered in light of the considerable difference that exists between the two groups as far as formal musical training is concerned. From that perspective, this difference may reflect the advantage provided by the formal training, rather than a qualitative difference in cognitive processes involved in music processing. From this interpretation, we can sketch the conclusion that the human brain is able to process music in a highly sophisticated manner that is only marginally enhanced by formal musical training: passive exposition to musical material provides enough stimulation to develop this abilities [374].

3.2 Makes a (big) difference

Expertise in specific domains raises the question of whether the special abilities can be generally transferred in other activities. In a recent study, Green & Bavelier suggest that experts in action video games exhibit enhanced abilities to sustain their attention on key stimulus features for an extended period of time [139]. Over the course of their training, musicians must acquire challenging skills and develop a strong ability to sustain attention on complex mixtures of sounds for a long period of time. Consequently, one might expect to observe a transfer of such sharp abilities onto broader perceptual

and cognitive tasks. A large body of studies have gathered evidences suggesting that musical training induces behavioral, functional and structural changes and that those changes modulate the processing of generic auditory processes [259]. The impact of musical training seems so important that it is now consensually considered as a model for brain plasticity [326, 409].

3.2.1 Anatomical differences

Playing a musical instrument require moving both hands, fingers, feet -and sometimes other unexpected parts of the body- in coordination. This intensive and daily motor activity is very likely to cause changes in the brain. Elbert *et al.* were the first to demonstrate that violinists players have enlarged cortical representation of the little finger of their left hand [108]. On the same trend, Schlaug *et al.* generalized this effect of motor training using morphometric MRI, as they published the first study that demonstrated that musicians possess a thicker *corpus callosum* -a structure of white matter bridging the two hemispheres-. This difference was interpreted to result from the repeated and intense sensory-motor information exchange between the right and left motor area during instrument playing [328]. Yet, these anatomical peculiarities concerned only musicians who had started their musical training before the age of seven, which may suggest that great brain plasticity may be required at the start of training to induce such changes. Further studies revealed that the asymmetry between the two *planum temporale* -a structure present on both hemispheres and known to be the siege of auditory information integration- was more pronounced in musicians with perfect pitch compared to non-musicians, as is appears in Figure 3.2.1 [327, 410]. Musicians with perfect pitch (also referred as absolute pitch) display a leftward asymmetry which could therefore reflect a prenatal disposition to develop absolute pitch and therefore, good musical abilities [177].

In a study using voxel-based morphometry (a technique that scopes the whole brain), Bermudez & Zatorre found a maximal difference between musicians

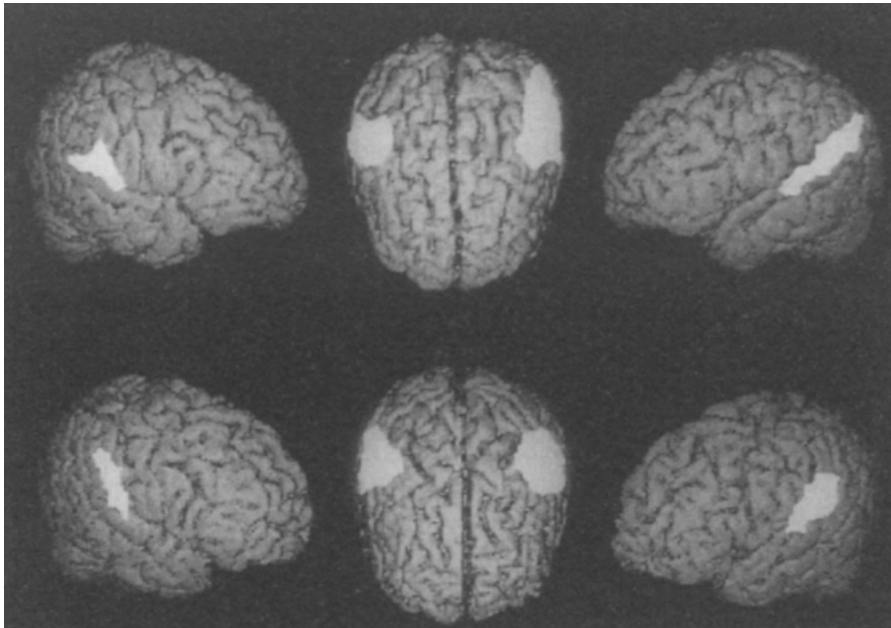


Figure 3.2.1: Brain surface projection of the right and left *planum temporale* in a musician (top panel) and a non-musician (bottom panel). From [327].

and non-musicians in grey matter density in a small region of the right *planum temporale*, a structure dedicated to pitch discrimination [19]. This increased grey matter was found in all types of musicians -independently of absolute pitch- thus suggesting that this anatomical differences are shaped by musical experience.

The Diffusion Tensor Imaging (DTI) technique provides a novel tool to explore the anatomical modulations caused by musical training. Indeed, this technique allows a fine tracking of axonal fibers in subcortical areas which can reveal increased organization of white matter tracks. Using DTI, Schmithorst & Milke found increased fiber density and orientation in the anterior *corpus callosum*, thus consistent with Schlaug *et al.* findings [329]. A subsequent study demonstrated that this increased white matter organization was correlated to the number of hours of musical practice, as displayed in Figure 3.2.2 [18].

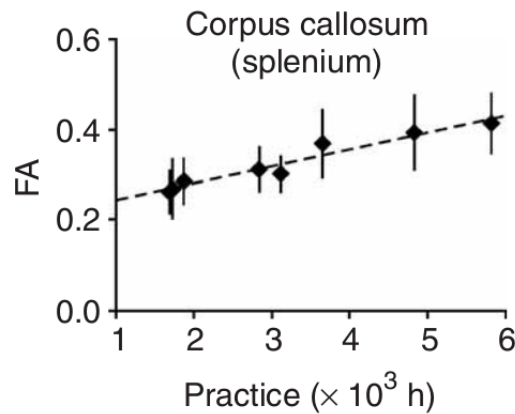


Figure 3.2.2: Correlations between participant's total amount of practice and white matter structure. Individual points display the mean FA (index of white matter organization) value across voxels for each participant. Vertical lines represent s.d. within that participant. Dashed lines are regression lines. From [18].

3.2.2 Behavioral differences

Pitch discrimination

As they constantly monitor the outcome -sound quality, tuning, timbre etc.- of their instruments, musicians potentially acquire very fine-grained pitch discrimination abilities. Indeed, numerous studies reported that musicians largely outperform non-musicians in frequency thresholds [182, 8, 249], especially for complex tones [249, 356]. In all the cited studies, the group differences were massive, raging from a factor of about two (in [182]) to a factor of six (in [249]), depending on the criteria for selecting musicians. It appears indeed that thresholds sizes are modulated by the number of years of musical training, as indicated by Figure 3.2.3 which displays the results collected on a cohort of musicians with varying degree of training duration [182].

However, although musicians exhibit a clear advantage on pitch discrimination tasks, it was also observed that the gap was rapidly filled with appropriate training [249, 182]. Micheyl et al. conducted two experiments on pitch

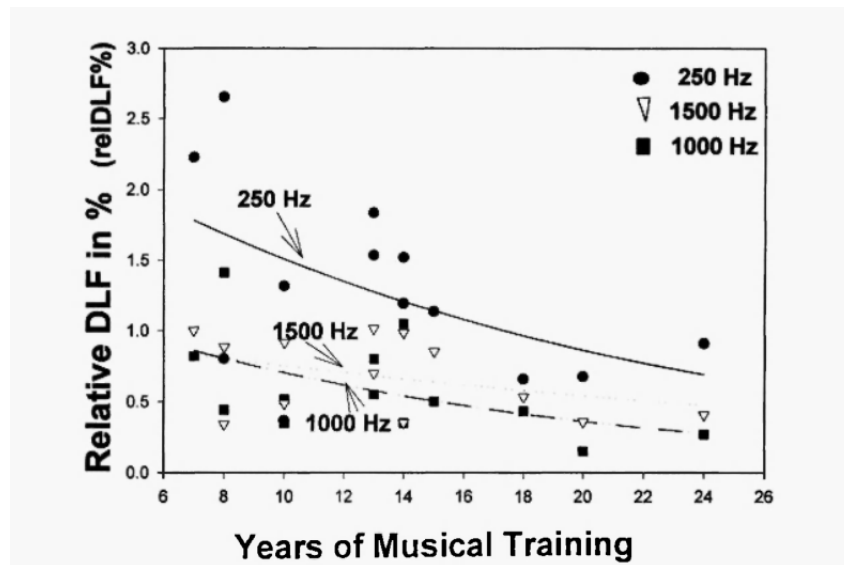


Figure 3.2.3: Points indicate averaged individual frequency threshold as a function of the number of years of musical practice for each reference frequency. Curves indicate best fitting exponential curves. From [182].

discrimination thresholds in musicians and non-musicians [249]. In the first experiment they observed large differences, consistently with previous findings. The second experiment was preceded by 14 hours of practice on pitch discrimination tasks. Results revealed that only 4 to 8 hours of practice in average was enough to wear off the group difference in thresholds.

Interval Categorization

Categorical perception (CP) corresponds to the fact that pairs of stimuli that lie along a physical continuum are perceived as belonging to distinct categories. As a consequence, perceptual discrimination thresholds between pairs of stimuli of same physical difference are smaller when they straddle a category boundary than when they belong to the same category. Thus CP can be explored through a combined methodology of identification and discrimination tasks. CP is considered as evidence that perceptual systems transform continuum sets of stimuli into discrete mental categories. Such

a perceptual phenomenon was observed in a speech contrast experiment conducted on infants [107].

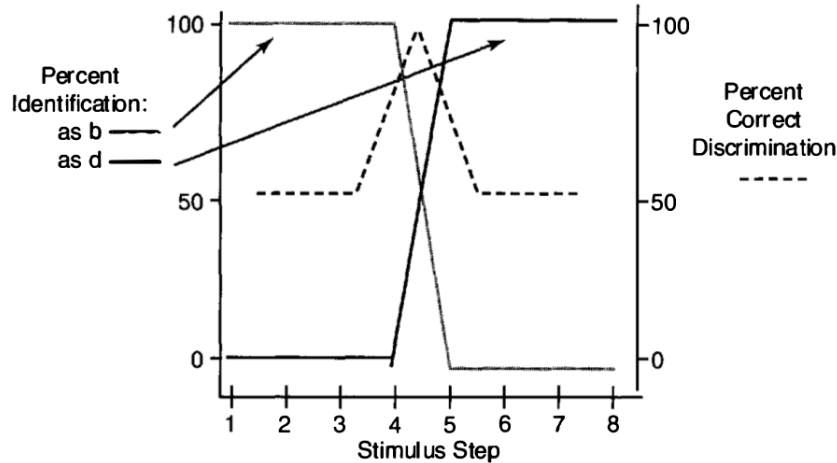


Figure 3.2.4: An idealized categorical speech perception function. The stimulus covers a continuous range and yet, perception dramatically shift from one category to the other at a particular point in the continuum. From [278].

Burns & Ward conducted an experiment on CP for intervals in musicians and non-musicians. Their paradigm consisted in varying an interval size by small increments between two endpoints, for instance between a minor third and a perfect fourth. This way, the pitch range encompasses three intervals: the minor third, the major third and the perfect fourth. As the identification task, listeners are presented the intervals in random order and are asked to assign them to one of the three categories. As a discrimination task, they are presented with pairs of intervals for a same-different judgment. Results exhibited evidence of CP for musicians but not for non-musicians which led them to claimed that CP of intervals for musicians were as robust as CP of phonetic items for non-musicians [54, 55].

CP induced by musical training constitutes an interesting case of a substantial alteration of abstract representation which alter perception itself and induced by long-term experience. However, it has received little attention in the past years, as few studies have tackled its neural counterparts. A behavioral

study was conducted in order to address a methodological default present in previous studies. Smith *et al.* sought to overcome the bias induced by formal training in musicians in giving novice a non-musical technique to perform the task [350]. Namely, participants were trained to identify intervals by matching them with the beginning notes of familiar tunes such as "What child is this?". Surprisingly, they found that the use of the alternative method enhanced performance of identification. However, in the task of discrimination, the alternative method did not contradict previous results, as the group effect was maintained. These results suggest that although CP measurements are sensitive to the degree to which the method require formal musical training, this cannot fully account for the difference observed between groups.

Encoding complex auditory signal

Concordant findings have also highlighted a positive correlation between musical training and acoustic signals encoding [261, 407, 218, 364] (for a review see [207]). This effect has been observed through the modulated response to complex acoustic signals in early areas of auditory processing (such as the cochlear nuclei in the brainstem and the inferior colliculus in the midbrain), thus fostering the view of auditory signal encoding as the result of a complex mixture of feed-forward and feed-back pathways [14] (see Figure 3.2.5) modulated by musicianship [363].

The efficiency, or accuracy of complex acoustic signals encoding in brainstem can be assessed by the adequacy between spectral and temporal features of the "neural code" referred as frequency-following response, observed in the subcortical structures and the those of the acoustic signal [59, 319]. The precision of the FFR response to speech has been observed be correlated to musical training, resulting a in a better tracking of the fundamental frequency of the voice (see Figure 3.2.6) [407] and a more synchronous neural responses to syllables [274]. The degree of accuracy for speech and intervals was

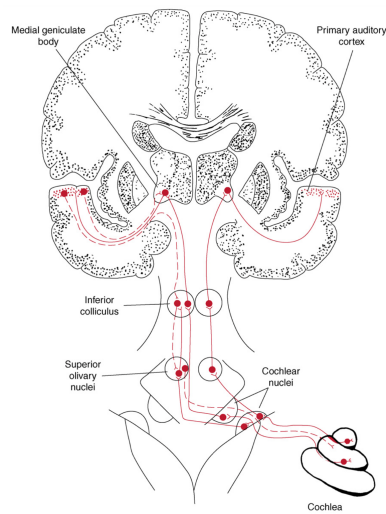


Figure 3.2.5: The hierarchical structure of auditory processing, from cochlea to primary auditory cortex. Red connections indicate bottom-up and top-down neural connections responsible for enhanced low-level encoding of complex signals for musically trained listeners. From [281, 181].

also shown to be modulated by the number of years of musical training [261, 407, 218, 364].

Further studies have gathered evidences suggesting that the effect of musicianship had also a positive impact on the comprehension of speech in noise [274, 273] and distorted speech signals [24] (although contradictory findings have been reported [317]). It was also reported that the deleterious effects of age on speech perception in noisy environment were attenuated by life-long musicianship [275]. The beneficial effects of musicianship were also reported in mistuning detection studies, in which a mistuned component is embedded in a complex harmonic tone [411]. Although it has been shown that aging causes a drop of mistuned component detection thresholds [142], musicianship seems to hinder this degradation [413, 412]. Although behavioral and ERP studies have repeatedly demonstrated that musicians and non-musicians only marginally differ in processing highly complex musical structures [33], recent findings have evidenced that musicians encode musical intervals [123], polyphonic melodies [124], sequential tones [384, 385]

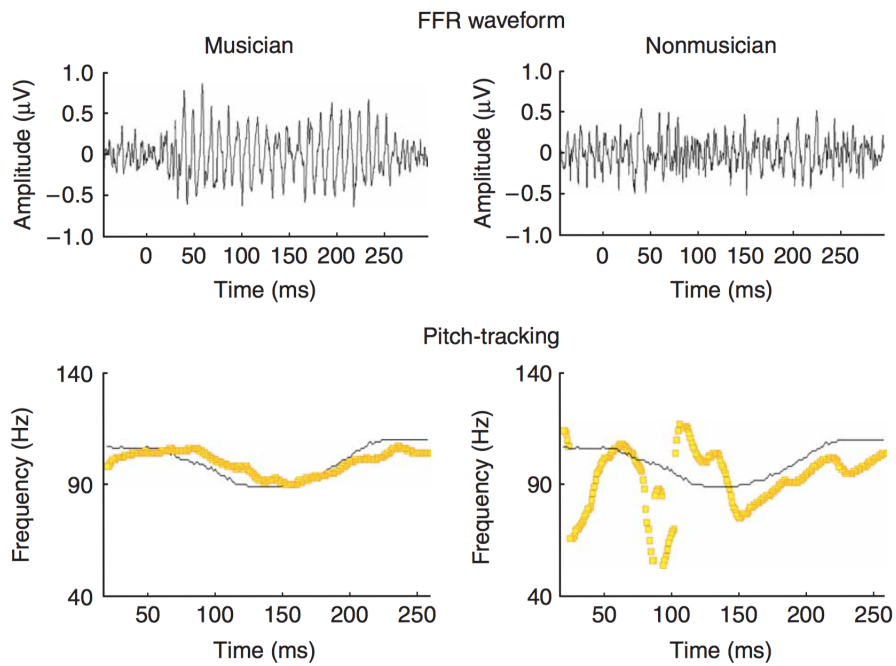


Figure 3.2.6: The top plots display FFR waveforms for musician and nonmusician listeners. The bottom plots displays the fundamental frequency of the speech signal as the (black line) and the FFR's primary periodicity (orange line) for musicians and nonmusicians. The tracking of fundamental frequency is more accurate for the musician listener. From [407].

and segregate spatially distributed sounds [263] with greater accuracy and precision than non-musicians.

Patel advocates for a causal effect of musicianship: musical training causes the enhanced neural encoding of speech sounds, a theory known as the well-named OPERA hypothesis [279, 280]. This hypothesis is consistent with findings suggesting that the advantage provided by musicianship does not pertain to enhanced sensory processes [317]. Rather, musical education modulates higher cognitive functions that in turn benefit auditory skills [363], which could therefore constitute a novel tool to develop auditory skills during childhood [207].

Conclusion

Since musical expertise originates from a rigorous daily training throughout life starting at a very young age, it plausibly induces changes in musicians' brain. Musical practice has been shown to induce anatomical differences in the *corpus callosum* and the *planum temporale* in musicians (for a review see [146]), although the extent to which those changes parallel particular musician traits, such as the possession of absolute pitch or the age of training starts, remain unclear [396].

These important brain modifications may therefore in turn modulate processing of acoustic signals [363], and among those, music. However, musical expertise seems to only marginally impact how complex musical structures are processed. Indeed, numerous studies have demonstrated that musicians and non-musicians in the Western culture only slightly differ in this area (for a review see [33]), especially considering their tremendous difference in musical education [206]. This has been hypothesized to result from the passive integration of the highly structured rules of Western music [374, 373].

But musicianship seems to enhance performance on other auditory processing tasks. Indeed, musical expertise has been reported to decrease pitch

discrimination thresholds [182], interval categorization [54, 55] and improve the encoding of musical [123, 124] and speech [281] signals. However, it should be noted that differences between musicians and non-musicians have sometime not been successfully replicated [317]. Moreover, it was also suggested that this difference can be rapidly overcome by training [249, 258, 269] or by adapting the task to non-musicians [350].

Overall, several elements point to a robust advantage of musicians in the processing of complex acoustic scenes [411, 124, 263], see also [296] and [368].

PART II : Experimental contribution

Chapter 4

The temporal dynamics of an auditory context effect

The work described in this Chapter is partially reported in: C. Chambers, S. Akram, V. Adam, C. Pelofi, M. Sahani, S. Shamma, and D. Pressnitzer. Prior context in audition informs binding and shapes simple features (under review). *Nature Communications*.

Abstract

In context effects, temporal parameters of contextual information, such as the durations of context presentation and the gap between context and test, have been demonstrated to induce different perceptual output in ambiguous stimuli [172]. In the Shepard paradigm (see detailed description in Section 1.2.5 on page 22) [342], it was found that a short sequence of tones heard prior to the ambiguous interval induced a strong assimilative bias: listeners tended to report the pitch-shift direction encompassing the frequency components of the context sequence [2]. The present series of experiment sought to characterize the temporal dynamics of this auditory context effect. To

this aim, the duration of the context was first varied, in order to investigate the minimal duration required to establish a significant effect. In a second experiment, a silent gap placed between the context and the test sequence was varied in duration, in order to investigate the remanence properties of the context effect. Finally, both parameters were varied, in order to observe interactions between the two factors.

Introduction

Contextual information plays a crucial role in everyday perception, as stimuli are never perceived in isolation. Rather, they belong to a complex network of sensory input and integrating prior knowledge, which interact to sketch a complex perceptual scene. Experimentally, context effects are studied by observing the influence of prior information (context) on a target stimulus and thereby contribute to explain how information integration occurs at different time-scale levels. Context effects can modulate the perception of non-ambiguous stimuli, creating for instance the illusion of a motion in a stationary image [11], but are especially well-suited for ambiguous stimuli, for their undetermined nature rises the sensitivity to the context [190].

As discussed earlier, the preceding context can yield to a contrastive effect or an assimilative effect. Contrastive effects have been reported in the visual modality in visual motion perception [11, 172], in the auditory modality in speech categorization [216, 231, 158, 160], and were even extended to other modalities [61, 414]. Assimilative effects, whereby the presentation of a context "attracts" the perception of a subsequent stimulus were also reported in vision [172] and in audition [131, 352, 63]. These two opposite context effects are believed to fulfill two essential goals of perceptual processes: contrastive effects may optimize information transfer and assimilative effects may subserve stabilization of perception over time [116, 336].

But what triggers the contrastive or assimilative nature of context effects? Temporal parameters, such as the duration of presentation of the context

and the gap separating context and test, seem to play a critical part in this regard. Scoping across different studies, it seems that prolonged presentation of context (in the order of seconds) induce contrastive effects, while short presentations (in the order of milliseconds) induce assimilative effects. In vision, an ambiguous motion was perceived as shifted in the opposite direction when preceded by a prolonged context [11] while much shorter context presentation have been reported to induce the exact opposite effect [286, 306]. In the auditory domain, comparable effects were found. Holt et al. reported a contrastive pitch perception following prolonged context presentation [158]. Namely, when preceded by what Holt et al. terms an "acoustic history" of several seconds, the perception of a subsequent vowel is shifted away from the frequencies present in the acoustic history, an effect then generalized to the perception of non-speech stimuli [158, 359, 362]. However, the synthesis of these concordant results is made difficult by the disparity of stimuli and methods. Besides, both in vision and in audition, other studies have reported contrastive effects following brief context presentations [305, 392, 6].

In order to assess directly the effect of temporal parameters, Kanai & Vesttraten conducted a study in which the duration of presentation and gap preceding an ambiguous motion was systematically varied [172]. Strikingly, they observe either contrastive or assimilative effects depending on the presentation duration and an interaction of the two temporal parameters (context and gap durations). They interpret these results as evidence for distinct adaptation mechanisms, both occurring in early stages of sensory processes but supported by neural populations with distinct temporal dynamics [152, 201, 200]. To date, not such opposite effects on the same stimulus were found in the auditory domain, but perhaps for lack of a systematic variation of the temporal parameters of the context.

Interestingly, Kanai et al. also report different temporal dynamics of the context effect depending on whether an ambiguous or unambiguous stimulus was used as context. Namely, they found that for the ambiguous context, the build up of an assimilative effect gradually increased and which did not decrease as the gap duration was prolonged. This slower temporal course

was interpreted as evidence of the higher-order nature of the context effect induced by the ambiguous stimulus. First, because it is caused by an ambiguous motion, the context effect cannot emerge from the excitation of early stage motion-tuning neurons, but rather, involves an higher level of representation. Second, the slow build-up of the effect could hardly be supported by adaptation of sensory neurons. In the auditory domain, a similar argument was made by Holt et al., when they found that context temporally separated from target elicited an increased context effect [158]. Indeed, while most context effects are interpreted as the result of early-stage adaptation processes, a growing body of evidence point to the existence of other types of context effects. In music cognition, the contribution of sensory vs. cognitive processes was revealed by an elegantly designed study, demonstrating that the cognition of chord sequences largely relies on high-level, cognitive processes [32]. In this perspective, the higher-order nature of these effects may be revealed through specific temporal dynamics [312]. Snyder et al. reported an assimilative effect in the perception of a "ABA" triplet: when the preceding context was perceived as integrated, the following sequence had increase chances to be perceived similarly [352]. This stabilization of perception over time may be the result of a statistical analysis of past, leading to a "regression to the mean", as it was recently reported in an auditory study on the perception of intervals [304].

The present study focuses on a recently observed context effect that biases the perception of an ambiguous pitch-shift. The ambiguous stimulus consist in the succession of two Shepard tones [342]. When two Shepard tones are separated by an interval of 6 semi-tones (st), it was consistently observed that listeners may report an upward or a downward shift with equal probability. Recently, it was discovered that a short sequence of tones presented before the ambiguous interval could induce a strong effect in the reported direction [2, 62] (online demonstration: <http://audition.ens.fr/dp/illusion/>), but the level at which this effect occurs in the auditory pathway remains unknown. To date, the observed effect of context is assimilative: listeners report the pitch-shift direction that encompasses the frequency region in

which context tones were placed. Three experiments were conducted with the aim to characterize the temporal course of this context effect. Systematically manipulating the durations of context presentation and gap, we could probe whether the assimilative nature of the context effect was dependent on temporal parameters and assess the dynamics of the effect in terms of establishment and decay.

4.1 Experiment 1.1

Expt 1.1 was conducted in order to investigate the minimal duration of a context sequence required to bias the perception of an ambiguous pitch-shift. To this aim, we presented listeners with the ambiguous interval preceded by a context tone of varying duration. At the end of the trial, listeners reported the pitch-shift direction of the final ambiguous interval. The experimental design was:

Subj*Context8

4.1.1 Method

Participants

Ten self-reported normal hearing participants (age: $M = 24.7$, $SD = 2.87$), eight women and two men, were included in this experiment after providing written informed consent. Six of them had never participated in an experiment involving Shepard tones. All were selected after the second version of the screening procedure (see below) and all were paid for their contribution.

Stimuli

SHEPARD TONES: All the tones used in this experiment were Shepard tones generated as described in previous experiments [63]. A Shepard tone consists

in nine pure tones, related by an octave relationship with respect to a base frequency F_b , added together to cover the audible range in frequency and amplitude-weighted by a fixed spectral envelope (see Figure 4.1.1). The spectral envelope was linear on the amplitude scale and logarithmic on the frequency scale ($M = 960$ Hz and $SD = 1 \log_2$ unit). The relative amplitudes of the components, $A(f)$, were computed using the equation below where f is the frequency of the component, cf , the central frequency of the envelope and σ , the standard deviation.

$$A(f) = \exp\left(\frac{1}{2} * \left(\frac{\log(f/cf)}{\log(2)/\sigma}\right)^2\right)$$

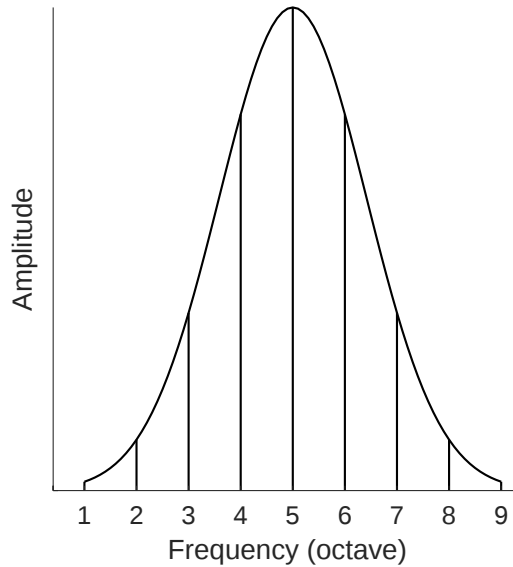


Figure 4.1.1: Shepard tones were generated by adding octave-related pure tones, weighted by a bell-shaped amplitude envelope.

A trial consisted in the succession of a context sequence composed of one context tone (C) followed by a test sequence composed of two test tones (T1-T2). The F_b for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz, to counterbalance possible idiosyncratic biases in pitch-shift direction preference [90, 63]. The interval between F_b s in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which

listeners typically report two pitch-shift directions with equal chances). The interval between Fbs in C and T1 was fixed at 3 st and 9 st so that C would induce the maximal bias toward an upward (3 st) or a downward (9 st) pitch-shift (see [2] and appendix D). The duration of the test tones was 125ms each, including 2.5ms raised-cosine onset and offset ramp. There was no delay between T1 and T2. The duration of the context tone varied from trial to trial between 0ms, 5ms, 10ms, 20ms, 40ms, 80ms, 160ms and 320ms, including a 2.5ms raised-cosine onset and offset ramp. There was no delay between C and T1. Each duration condition was repeated forty times and presentation order was scrambled within each test session. Figure 4.1.2 displays a schematic representation of one trial.

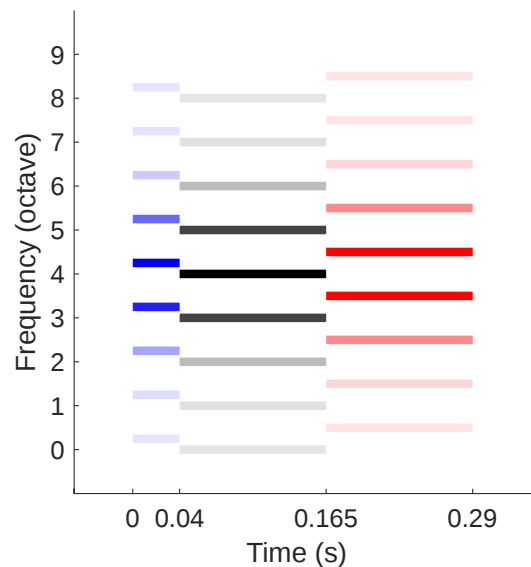


Figure 4.1.2: In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is, that corresponds here to an "upward" pitch-shift.

INTER-TRIAL TONES: Pilot data revealed that an across-trial effect could occur, by which test tones of the previous trial biased the perception of the test tones of the subsequent bias, thus possibly interacting with the

effect induced by very short context tones. In order to minimize this effect, participants were played an inter-trial sequence between trials. Inter-trial tones were designed to "saturate" a broad range of frequency and thus "clean up the ears" between each trial. Therefore, the inter-trial tones were similar to Shepard tones but with a half-octave relationship between tones. The base frequency F_b of each inter-trial tones was randomly drawn uniformly between 60 Hz and 120 Hz. Each inter-trial tone had a duration of 125ms, including a 5ms raised-cosine onset and offset ramp. The inter-trial sequence consisted in the succession of five inter-trial tones separated by a short silence of 125ms.

Screening procedure

The task required in this experiment requires participant to report the pitch-shift direction resulting from an interval. For this task, individual display different performances [339]. As a screening procedure we therefore selected participants on their abilities to accurately report pitch-shift directions in the basis of their score in a screening test that was carried out systematically prior to participation.

In this test, participants had to report the pitch-shift direction of an interval T1-T2 either composed of two pure tones, or composed of two Shepard tones. For the pure tones intervals, the frequency of T1 was randomly drawn between 960Hz and 1920Hz. The frequency of T2 was shifted by an interval of 3 st, 2 st or 1 st. For the Shepard tones intervals, the two Shepard tones were designed as previously described ([Stimuli](#)). The F_b of T1 was randomly drawn in a frequency range of 60 to 120Hz. The F_b of T2 was obtained by shifting F_b of T1 by an interval of 3 st, 2 st or 1 st. The different intervals between T1 and T2 corresponded to distinct levels of difficulty, 3 st being the easiest condition and 1 st the most difficult [339].

The duration of each tone was 125ms, including a 5ms onset and offset ramp. T1 and T2 were played separated by a delay of 125ms. The order

presentation between T1 and T2 was randomized so that both upward and downward shifts were played. After each interval, participants had to report their response by pressing the 1 or 2 touch on the keyboard. Through the display of a green or red patch on the screen, listeners were provided with feedback on their response.

In a first version of the screening test, all difficulty conditions (3 st, 2 st and 1 st) were presented in random order with 10 repeats per condition in each session. Participants who obtained a minimum score of 80% correct for the 1 st interval for both Shepard tones and pure tones in one session were selected. This version of the test was used to select participants in Expmt 1.2 (which was, chronologically speaking, the first experiment that was conducted). Yet, as version 1 excluded an excessively large proportion of participant, a second version of the screening test was designed.

In the second version of the screening test, which was used for all the other experiments, the difficulty level (i.e. interval size) was fixed in blocks containing each 40 trials. The 3 st interval was presented first, followed by the 2 st and 1 st conditions. This way, participants had the opportunity to become familiarized with the difficulty of the task before moving the next difficulty level. If a participant obtained a score above 80% in one session, the next session moved on to the next difficulty level, until the score of 80% was reached for the 1 st difficulty level. If a participant obtained a score below 80%, after three blocks of the same difficulty condition, the screening procedure stopped and the participant was excluded.

Apparatus and procedure

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustic Company) during one session containing 320 trials. The total duration of the test session was approximately forty minutes. Stimuli were played diotically through an RME fireface 800 sound-card at a 16-bit resolution and a 44.1 kHz sampling rate. They were presented to both

ears simultaneously through Sennheiser HD 250 Linear II headphones. The presentation level was 65 dB SPL, A-weighted. At the end of each trial, listeners were asked to report whether the interval formed by the two final tones was going upward or downward in pitch. To give their answer, they had to press either key 1 (downward shift) or key 2 (upward shift) of the keyboard. The inter-trial sequence was then played.

Data analysis

For each listener and context duration condition, the proportion of "up" $P(\text{up})$ responses was computed from the pitch-shift direction responses. Since this experiment focused on the bias of the context depending on its duration a more synthesized measure of the context was computed for each listener and each context duration. The proportion of "biased" responses $P(\text{bias})$ was indeed obtained by computing the proportion of time listeners responded with a bias in the direction expected from previous findings, that is with a pitch-shift encompassing the frequency region where the frequency components of the context tone were [62, 2]. This way, $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones. A $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift. An absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5.

In order to control for across trial-effects, a complementary analysis was conducted. We wanted to make sure that in the control condition -i.e. in the absence of a context tone- T_2 of the previous trial $T_2(\text{prev})$ was not biasing the perception of the current trial T_1 - $T_2(\text{curr})$. To this aim a $P(\text{bias})$ was computed similarly as described above for each control condition with $T_2(\text{prev})$ as the context tone.

4.1.2 Results

Pitch-shift direction response

The results are displayed in Figure 4.1.3: the $P(\text{bias})$ averaged across participants is plotted as a function of the context tone duration. Thin lines represent individual results. As expected, in the control condition with no context tone, the $P(\text{bias})$ was near 0.5. As the duration of the context increases, the $P(\text{bias})$ builded up gradually. A one-way repeated measure ANOVA conducted on the context duration as within subject factor confirmed that the duration of the context had a significant effect on the $P(\text{bias})$ ($F(7, 63) = 30.47, p = 2e - 16$).

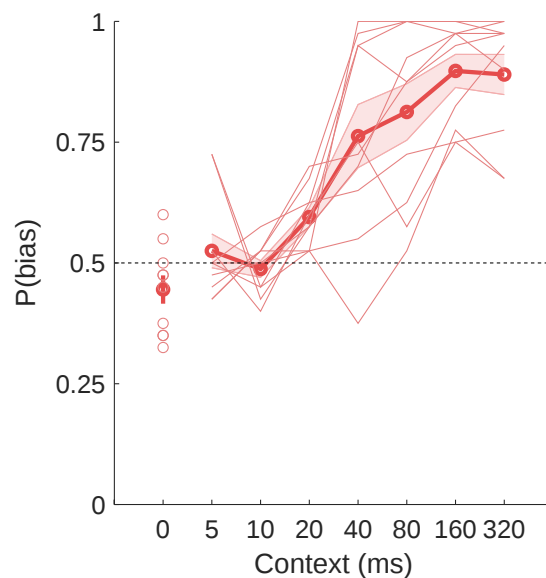


Figure 4.1.3: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean and lines indicate individual results.

The figure indicates that the assimilative effect builds up very quickly as it appears for a context sequence as short a 20ms. This was confirmed by a series of corrected t -tests performed on the context bias measure using a Bonferroni correction (see Table 5.1 for all t -test results). A sequence of 20ms was sufficient to induce an effect significant bias on the direction of

Context (ms)	Mean P(bias) (SE)	t -test(df = 9)	p -value (Bonferroni corrected)
0	0.44 (0.03)	-1.8773	0.72
5	0.52 (0.03)	0.7151	3.93
10	0.48 (0.02)	-0.7276	3.88
20	0.59 (0.02)	4.8347	0.007**
40	0.76 (0.06)	4.0118	0.024*
80	0.81 (0.06)	5.3435	0.003**
160	0.89 (0.03)	11.6066	8.1e-06***
320	0.89 (0.04)	9.3901	4.8e-05***

Table 4.1: Post-hoc analysis results are reported. T-tests with a Bonferroni correction were conducted and the proportion of "biased" response P(bias) was compared to 0.5 in each context condition. Reported p -values are multiplied by the number of observations $N = 8$.

the ambiguous test interval (one-sample t -test, $t(9) = 4.83$, $p = 0.007$, the p -values reported here are multiplied by $N = 8$).

Across-trial effect

Since the context tone was very short in certain condition and given that the effect revealed to be extremely sensitive, a complementary analysis was carried out to ensure that the perception of a current trial was not biased by the previous trial. To this aim, a P(bias) was computed for each listener for all control condition trials (trials with no context tone) with T2(prev) as the context tone, as described above. The P(bias) computed for each participant did not significantly differed from a neutral bias (one-sample t -test: $M = 0.5$, $SD = 0.07$, $t(9) = 0.32$, $p = 0.76$).

4.1.3 Summary

The experiment was conducted to determine the minimal duration of a context tone to induce a significant bias on the ambiguous interval. To this aim,

context sequences of various duration preceded the ambiguous interval and the pitch-shift direction reported from the interval constituted a measure of the direction and strength of the bias for the different context conditions. The context strength increased as the context duration increased from 5ms to 20ms and reached maximum for a context duration of 160ms. The context effect could be established extremely rapidly since a context tone of only 20ms was sufficient to induce a significant bias. Besides, as previously observed, the perceived pitch-shift "cut through" the frequency region in which the context tone was placed, relative to the interval for all tested duration conditions.

4.2 Experiment 1.2

The aim of Exp. 1.2 was to explore another temporal aspect of the context effect: its remanence, that is, the duration for which the effect of the context was maintained. In each trial, a test sequence of an ambiguous interval of Shepard tones was preceded by a context sequence. Between the context sequence and the test sequence, a gap of various duration was introduced during which participants were required to stay passive and wait for the test tones to be played. At the end of each trial, listeners had to report the pitch-shift direction of the final ambiguous interval. The experimental design was:

Subj*Gap8

4.2.1 Method

Participants

Ten self-reported normal-hearing listeners ($M = 26.3$, $SD = 1.54$), four men and six women were included in this experiment after providing written

informed consent. Two of them had never participated in an experiment involving Shepard tones. All were paid for their contribution. All were selected after the first version of the [Screening procedure](#) and all were paid for their contribution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described above ([Stimuli](#)). Each trial consisted in the succession of a context sequence of five context tones (Cs) and a test sequence composed of two test tones (T1-T2). The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz. The interval between Fbs in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which listeners typically report two pitch-shift directions with equal chances). In each trial, all the Fb of the five context tones were randomly drawn in a frequency range of either 0 st - 6 st or 6 st - 12 st from the Fb of T1. This way, in half of the trials the context sequence would induce an upward shift (0 st - 6 st) and in the other half, a downward shift (6 st - 12 st), consistently with previous findings [2]. The duration of the test tones was 125ms each, including 5 ms raised-cosine onset and offset ramp. There was a delay of 125 ms between T1 and T2.

The duration of the context tone was 125 ms, including a 5 ms onset and offset ramp. Each context tone was separated by a delay of 125 ms. A silent gap was introduced between the context sequence and the test sequence C-T1 which duration varied between 0.5, 1, 2, 4, 8, 16, 32 and 64 s. Each gap duration condition was repeated twenty times and the presentation of conditions was shuffled. Those temporal parameters were chosen consistently with previous findings (see Appendix D) so that the context sequence followed by the minimal gap duration would elicit a strong context. Figure 4.2.1 displays a schematic representation of one trial. There were no inter-trial sequence played between trials.

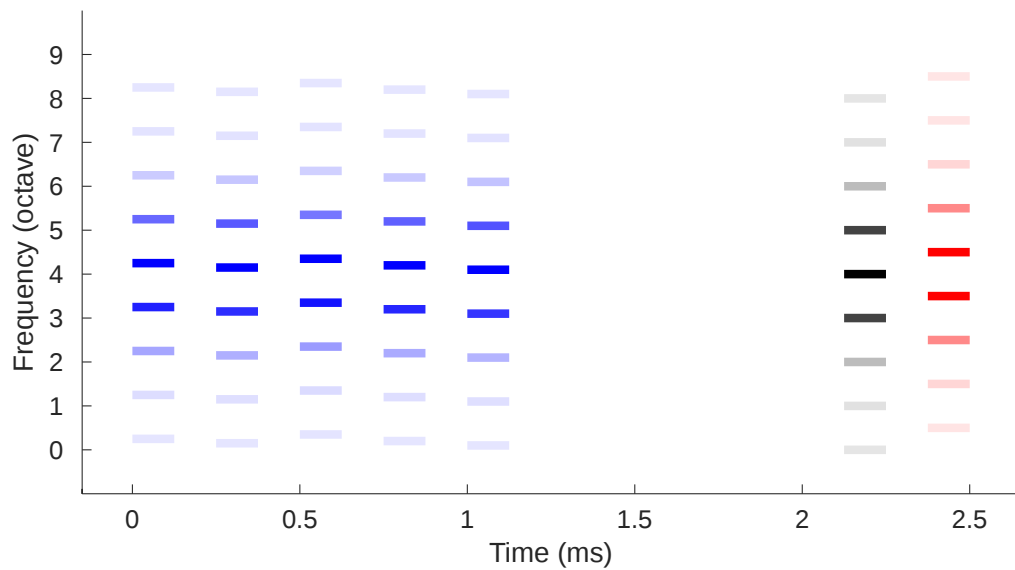


Figure 4.2.1: In a trial, a context sequence composed of five tones preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. The duration of the silent gap between the context sequence was varied.

Apparatus and procedure

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustic Company) during two sessions containing 80 trials and lasting thirty minutes each. Listeners took a short break between the two sessions. The total duration of the experiment was approximately one hour. Stimuli were played diotically through an RME fireface 800 sound-card at a 16-bit resolution and a 44.1 kHz sampling rate. They were presented to both ears simultaneously through Sennheiser HD 250 Linear II headphones. The presentation level was an above threshold of 65Hz. At the end of each trial, listeners were asked to report whether the interval formed by the two final tones was going upward or downward in pitch. To give their answer, they had to press either key 1 (downward shift) or key 2 (upward shift) of the keyboard.

Data analysis

The proportion of "up" response $P(\text{up})$ was computed from the pitch-shift direction responses for all listeners and each gap duration condition. As previously described, a proportion of "biased" responses $P(\text{bias})$ was computed from the $P(\text{up})$. A $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones i.e. assimilative bias, whereas $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift, i.e. contrastive bias ; an absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5. In order to control for across trial-effects, the effect of $T2(\text{prev})$ was computed for each listeners and all conditions, as described above.

4.2.2 Results

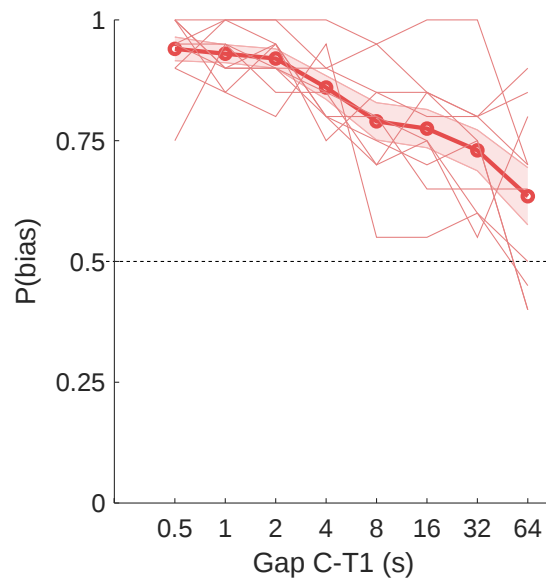


Figure 4.2.2: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean and lines indicate individual results.

Pitch-shift direction response

The results are displayed in Figure 4.2.2. The $P(\text{bias})$ averaged across participants is plotted as a function of the gap C-T1 duration. Thin lines represent individual results. Consistently with previous results, the $P(\text{bias})$ was near maximum for the shortest C-T1 condition, indicating a strong effect of the context sequence. As the C-T1 duration increased, the effect gradually decayed, yet it was not neutral even for the longest C-T1 duration. A one-way ANOVA with gap C-T1 as a within factor confirmed that it has a significant effect on the $P(\text{bias})$ ($F(6, 63) = 14.48$, $p = 4.5e - 11$). To investigate at which gap duration the effect of the context was significant, we conducted a series of t -tests with a Bonferroni correction which details are reported in Table XX. This results revealed that the context sequence induced a bias significantly different from neutral (0.5) even after a 32 s silent gap ($t(9) = 5.43$, $p = 0.003$).

Gap (s)	Mean P(bias) (SE)	<i>t</i> -test(df = 9)	<i>p</i> -value (Bonferroni corrected)
0.5	0.94 (0.02)	17.962	1.8e-07***
1	0.93 (0.02)	23.169	1.9e-08***
2	0.92 (0.02)	21	5.90E-09***
4	0.86 (0.02)	14.696	4.72e-08***
8	0.79 (0.04)	7.515	0.0002***
16	0.77 (0.04)	6.941	0.0005***
32	0.73 (0.04)	5.438	0.003**
64	0.63 (0.06)	2.281	0.38

Table 4.2: Post-hoc analysis results are reported. *t*-tests with a Bonferroni correction were conducted and the proportion of "biased" response P(bias) was compared to 0.5 in each gap condition. Reported *p*-values are multiplied by the number of observations $N = 8$.

Across-trial effect

To ensure that no across-trial effect impacted the results, the analysis of the previous trial effect was conducted. A P(bias) was computed for each listener for all conditions trials with T2(prev) as the context tone. The P(bias) computed for each participant did not significantly differed from a neutral bias (one-sample *t*-test: $M = 0.48$, $SD = 0.08$, $t(9) = -0.54$, $p = 0.59$).

4.2.3 Summary

Exp. 1.2 was conducted to investigate the remanence properties of a context effect on the perception of the ambiguous interval. A context sequence containing five tones and inducing a strong bias was placed before the ambiguous interval. A silent gap of various durations was introduced between the two sequences. The pitch-shift direction reported from the interval constituted a measure of the direction and strength of the bias for the different gap duration conditions. The results suggest that a context effect was still significant after a silent pause as long as 32 seconds, although individual

results suggest that for some listeners, the context effect is still quite strong after over one minute of silent gap. Indeed, individual lines reveal that for some listeners, the effect of the bias was still maximally strong, even after 32 seconds. Since we could not control the behavior of participants during this long silent pause, we can suppose that some listeners adopted strategies for memorizing the context sequence, leading to enhance its effect. This point will be examined in the discussion section.

4.3 Experiment 1.3

The last experiment presented in this chapter was conducted in order to investigate the interaction between the duration of the context and the duration of the silent gap separating the context and the ambiguous interval. To this aim, the duration of the context and the silent gap separating the context sequence and the test were both varied. At the end of each trial, listeners had to report the pitch- shift direction of the final ambiguous interval. As a control experiment, an additional task was carried out, in which participants were presented either no context or a very short context before the test tones. After providing their pitch-shift response, they had to report whether they heard a brief tone before the test interval. In the main experiment the design was:

Subj*Context4*Gap6

4.3.1 Method

Participants

Ten self-reported normal-hearing listeners ($M = 24.6$, $SD = 3.54$), two men and eight women participated in the experiment. All of them successfully passed through the screening test (version 2, see [Screening procedure](#)).

Among them, two had never participated in a previous experiment involving Shepard tones. All were paid for their contribution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described above (Stimuli). Each trial consisted in the succession of a context sequence of one tone (C) and a test sequence composed of two test tones (T1-T2). The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz.

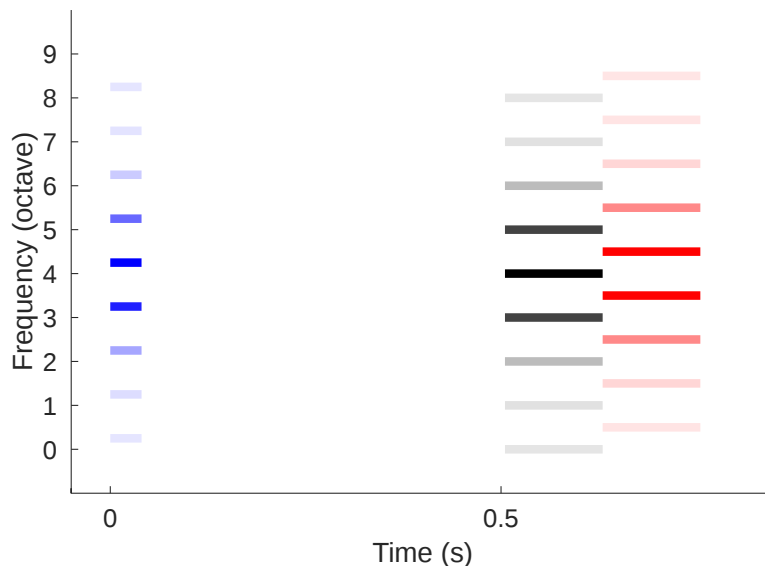


Figure 4.3.1: In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. The duration of the tone and the silent gap between the context sequence were varied.

The interval between Fbs in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which listeners typically report two pitch-shift directions with equal chances). The interval between Fbs in C and T1 was

fixed at 3 st and 9 st so that C would induce the maximal bias toward an upward (3 st) or a downward (9 st) pitch-shift. The duration of the test tones was 125ms each, including 2.5 ms raised-cosine onset and offset ramp. There was no delay between T1 and T2. The duration of C was varied between 20, 40, 80 and 160 ms. A silent gap was introduced between the context tone and the test tones. Its duration varied between 0, 0.25, 0.5, 1, 2 and 4 s. Each gap and context duration combination was repeated forty times and the presentation order was shuffled within test sessions. Figure 4.3.1 displays a schematic representation of one trial.

In order to prevent across-trial effects, an inter-trial sequence of five inter-trial tones was played, as previously described [Stimuli](#). Each inter-trial tone had a duration of 125 ms, including a 5ms raised-cosine onset and offset ramp and was separated from the subsequent tone by a short silence of 125 ms.

Additionally, every participant completed the control experiment. Its aim was to control that the very short context tone immediately followed by the test tones was not masked by a backward masking phenomenon [\[297\]](#). During the control experiment, listeners were presented a test sequence immediately preceded by a very short context tone of 20 ms or no context tone at all. The test tones and context tone were designed as described in the main experiment. The two context conditions were presented forty times and shuffled within test session. Between each trial, an inter-trial sequence was played, similar to previously described.

Apparatus and procedure

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustic Company) during three sessions. The main experiment was composed of two sessions of 480 trials of 50 minutes each. Then, listeners completed the control condition in one session of 80 trials which lasted 10 minutes. The total duration of the experimental session was approximately

two hours and between each sessions listeners took a short break. Stimuli were played diotically through an RME fireface 800 sound-card at a 16-bit resolution and a 44.1 kHz sampling rate. They were presented to both ears simultaneously through Sennheiser HD 250 Linear II headphones. The presentation level was 65 dB SPL, A-weighted. At the end of each trial of the main experiment, listeners were asked to report whether the interval formed by the two final tones was going upward or downward in pitch. To give their answer, they had to press either key 1 (downward shift) or key 2 (upward shift) of the keyboard. The inter-trial sequence was then played. During the control experiment, listeners had to report the pitch-shift direction, as in the main experiment. Then, they were asked to report whether they heard c tone before the final tone pair. To give their answer, they had to press either key 1 or key 2 of the keyboard. A green or red box appeared on the screen after each detection response to provide feedback. The inter-trial sequence was then played.

Data analysis

The proportion of "up" response $P(\text{up})$ was computed from the pitch-shift direction responses for all listeners and each gap duration condition. As previously described, a proportion of "biased" responses $P(\text{bias})$ was computed from the $P(\text{up})$. A $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones, whereas $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift ; an absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5. This analysis was carried out for the main experiment and the control experiment. Additionally, in the control experiment a performance score was computed for each listeners. It corresponded to the proportion of correct response in the detection task $P(\text{hit})$.

4.3.2 Results

Pitch-shift direction response

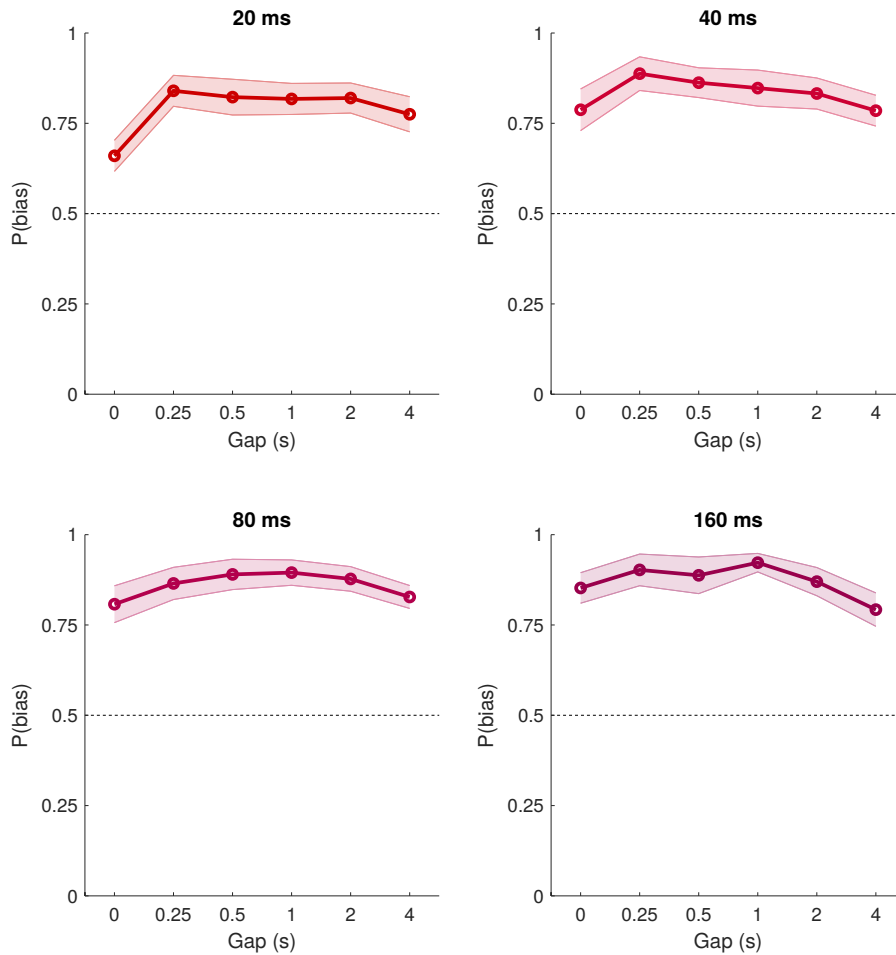


Figure 4.3.2: Each panel displays the proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners for one context duration. Shaded areas indicate ± 1 standard error about the mean.

The results are displayed in Figure 4.3.2. The $P(\text{bias})$ averaged across participants is plotted as a function of the gap C-T1 duration for each context

duration conditions in separated panels. As expected, the $P(\text{bias})$ tends to decay with the gap duration increasing and this effect is more pronounced for shorter context durations (20 ms and 40 ms). Remarkably, the $P(\text{bias})$ quickly build up as the gap duration increases. It was mostly pronounced for very short context durations (20 ms and 40 ms) in the two upper panels.

A two-way repeated measure ANOVA with gap and context durations as two within subject factors was conducted. It confirmed that both factors had a significant effect on $P(\text{bias})$ (context: $F(3, 27) = 15.84, p = 3.8e - 06$ and gap: $F(5, 45) = 11.84, p = 2.3e - 07$).

Interaction between context and gap duration

The interaction between the two factors was also significant ($F(15, 135) = 2.21, p = 0.008$). To investigate what pairs of conditions of gap and context duration were responsible for the interaction, a post-hoc multiple comparison (Bonferroni correction) was run. The results are reported in Table 4.3. Only pairs of conditions with a significant difference are reported.

This post-hoc analysis revealed that all the pairs of gap/context duration that significantly differed contained the combination of condition "gap 0s - context 20ms". This suggests that the duration of the silent gap has a specific effect for the short context tone duration: for this condition, the silent gap enhances the context bias. This interaction effect corresponds to the "build-up" that is observed in the first upper panel in Figure XX.

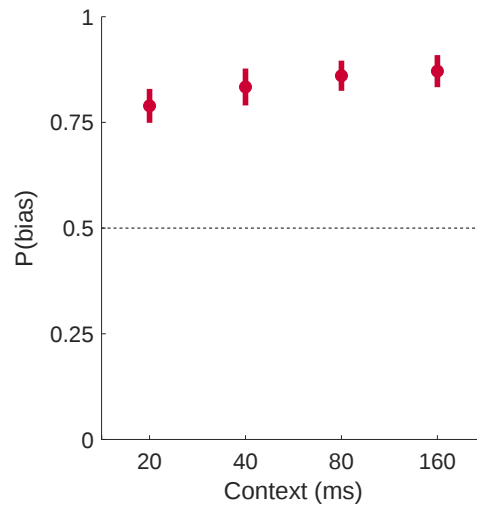
To further detail the results, we then investigated the effect of each factor, context and gap durations, independently. A $P(\text{bias})$ was computed plotted for each context duration condition in Figure XX and for each gap condition in Figure XX. A post-hoc multiple comparison (Bonferroni correction) analysis was conducted on each factor independently.

Pair of conditions	Z-value	Pr(> z)
gap2.dur1 - gap1.dur1	6.11	2.62e-07 ***
gap3.dur1 - gap1.dur1	5.523	9.19e-06 ***
gap4.dur1 - gap1.dur1	5.35	2.38e-05 ***
gap5.dur1 - gap1.dur1	5.43	1.49e-05 ***
gap6.dur1 - gap1.dur1	3.90	0.025 *
gap1.dur2 - gap1.dur1	4.33	0.004 **
gap2.dur2 - gap1.dur1	7.73	2.94e-12 ***
gap3.dur2 - gap1.dur1	6.88	1.62e-09 ***
gap4.dur2 - gap1.dur1	6.37	5.12e-08 ***
gap5.dur2 - gap1.dur1	5.86	1.25e-06 ***
gap6.dur2 - gap1.dur1	4.24	0.005 **
gap1.dur3 - gap1.dur1	5.01	0.0001 ***
gap2.dur3 - gap1.dur1	6.96	8.89e-10 ***
gap3.dur3 - gap1.dur1	7.81	1.47e-12 ***
gap4.dur3 - gap1.dur1	7.98	3.68e-13 ***
gap5.dur3 - gap1.dur1	7.39	3.98e-11 ***
gap6.dur3 - gap1.dur1	5.69	3.44e-06 ***
gap1.dur4 - gap1.dur1	6.54	1.67e-08 ***
gap2.dur4 - gap1.dur1	8.24	6.13e-14 ***
gap3.dur4 - gap1.dur1	7.73	2.94e-12 ***
gap4.dur4 - gap1.dur1	8.92	2e-16 ***
gap5.dur4 - gap1.dur1	7.13	2.62e-10 ***
gap6.dur4 - gap1.dur1	4.5	0.001 **

Table 4.3: Post-hoc multiple comparisons of means for pairs of gap and context duration factor. Only pairs with a significant difference are reported.

Context duration

Figure 4.3.3 report the results for the context duration. In Figure 4.3.3 each bar corresponds to the mean and SEM of $P(\text{bias})$ for one context duration condition averaged across participants. In table XX, only the pairs of condition with a significant difference are reported.



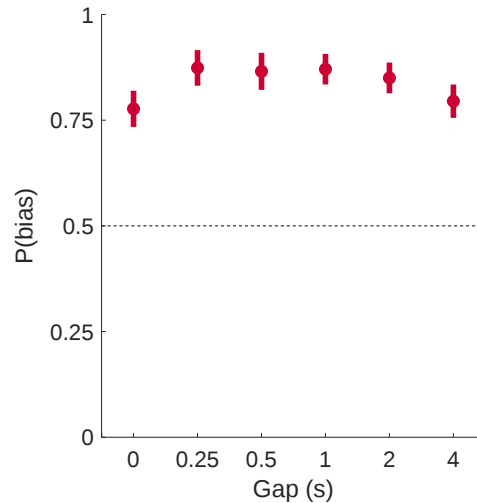
Pair of conditions	Z-value	Pr(> z)
dur2 - dur1	3.43	0.003 **
dur3 - dur1	5.48	2.49e-07 ***
dur4 - dur1	6.31	1.59e-09 ***
dur4 - dur2	2.88	0.0234 *

Figure 4.3.3: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Bars indicate ± 1 standard error about the mean. Post-hoc multiple comparisons of means for pairs of context duration factor. Only pairs with a significant difference are reported.

Gap duration

Figure 4.3.4 report the results for the gap duration. In Figure 4.3.4 each bar corresponds to the mean and SEM of $P(\text{bias})$ for one gap duration condition

averaged across participants. In table 4.3.4, only the pairs of condition with a significant difference are reported.



Pair of conditions	Z-value	Pr(> z)
gap2 - gap1	5.62	2.87e-07 ***
gap3 - gap1	5.14	3.94e-06 ***
gap4 - gap1	5.43	8.07e-07 ***
gap5 - gap1	4.24	3.33e-04 ***
gap6 - gap2	-4.56	7.38e-05 ***
gap6 - gap3	-4.09	6.28e-04 ***
gap6 - gap4	-4.38	1.72e-04 ***
gap6 - gap5	-3.19	0.021 *

Figure 4.3.4: Top panel: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each gap duration and averaged across listeners. Bars indicate ± 1 standard error about the mean. Bottom panel: Post-hoc multiple comparisons of means for pairs of gap duration factor. Only pairs with a significant difference are reported.

Time from onset

It is also possible that the most impacting factor is the duration between the onset of C and T1, which corresponds to a combination of the context and

the gap duration, hence not visible when plotting the $P(\text{bias})$ as a function of gap and context durations. To visualize the effect of this delay between C and T1 onset, the onsets are computed from each combination of gap and context durations. In Figure 4.3.5, the $P(\text{bias})$ is plotted as a function of the C-T1 onset duration on a real scale. Only the first onset durations are plotted since the effect rapidly reached a maximum plateau. An ANOVA confirmed that the effect of onset was significant ($F(23, 207) = 6.728, p = 2.66e-15$).

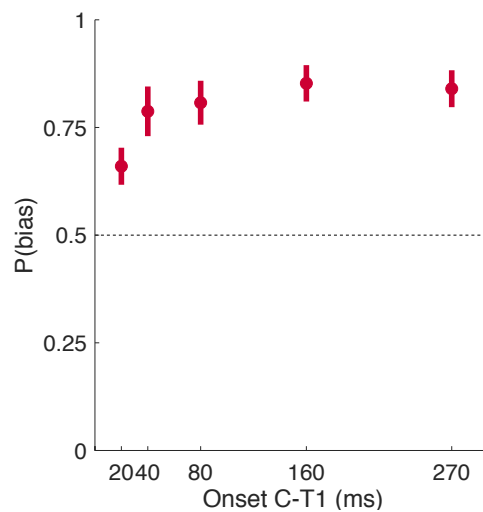


Figure 4.3.5: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each C-T1 onset delays and averaged across listeners in real scale. Bars indicate ± 1 standard error about the mean.

Control experiment

A performance score at the detection task was computed as the percentage of correct responses for each listener. The performance was very high (percent correct: $M = 0.97, SD = 0.02$). Listeners had no difficulty detecting to context tone of 20ms. Figure 4.3.6 displays the $P(\text{bias})$ averaged across participants for the two context condition (no context and 20 ms context). Individual dots correspond to individual results. Consistently with previous results, $P(\text{bias})$ was not neutral (0.5) for the 20 ms context condition

(one-sample t -test, $t(9) = 3.4$, $p = 0.007$). A pairwise t -test revealed that the context bias for the two conditions significantly differed ($t(9) = -3.3$, $p = 0.009$).

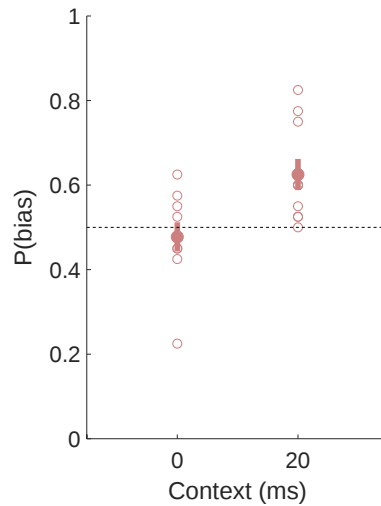


Figure 4.3.6: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Bars indicate ± 1 standard error about the mean.

4.3.3 Summary

Exp. 1.3 was conducted to investigate the interactions between the context duration and the silent gap duration. Therefore, both temporal parameters were varied as within subject factors and their effect on the proportion of “biased” response was observed. The results were consistent with previously presented results in the same chapter, namely, the context duration was critical to establish the context effect and this build up occurred extremely rapidly, since a context tone of 20ms, with no silent gap between the context sequence and the test sequence was enough to induce a significant effect. A control experiment was conducted to ensure that the absence of effect for tone duration shorter than 20ms was not due to a backward masking effect. Very high detection performance on this task confirmed that listeners were

indeed able to detect the very short tone. Consistently with Exp. 1.2, after a silent gap of 4s, the effect remained very strong. The most remarkable results lies in the fact that for short gaps and short context durations, an interaction between the two parameters occurred, whereby the build up of the effect was much more pronounced when no silent gap separated the two sequences, suggesting that the silent gap had a more critical role for very short context durations than for longer ones.

4.4 Discussion

Three experiments were conducted in order to characterize the temporal dynamics of a recently observed context effect on the perception of an ambiguous pitch-shift. Indeed, it was shown that when a short sequence of five tones preceded an interval of two Shepard tones separated by 6 st, the perception was strongly biased in an assimilative manner: the reported pitch-shift encompasses the frequency region in which the context tones were placed [2].

We observed extremely fast establishment properties (20 ms). Such a rapid time-course is outstanding in the literature. In vision, it was found that the exposure to an energy-based motion of 80 ms could induce a priming effect on a subsequent ambiguous motion [286, 172]. In audition, short presentations of frequency glides were shown to bias the perception of subsequent stimuli. These effects were observed for context presentation durations as short as 100ms [6, 392]. In the mentioned studies, shorter durations of the context were not tested, it is therefore impossible to assess whether even shorter context, as short as what we have tested in Exp. 1.1, could elicit a similarly strong bias. A similar "logarithmic build-up" was reported in visual aftereffects [312], but following longer context presentations. Hence, this constitutes the first piece of evidence for extremely rapid form of context effect that bias the perception of an ambiguous interval.

Interestingly, all tested context durations induced a context effect that was assimilative. In previous studies, it was found that varying the context duration could lead to opposite context effects; short presentations would induce assimilative and long presentations would induce contrastive effects for the same tested stimulus [172, 45]. These results were interpreted as reflecting distinct processes of neural adaptation with distinct time-scales [44], a phenomenon that could also arise from neurons in the auditory pathway [382]. Here the context effect observed was assimilative for all the tested conditions, which suggests that it may result from a single mechanism that is triggered by very short exposition durations and persist for very long exposition durations.

In Exp. 1.2 we observed a very slow decay of the context effect: it was still significant after a silent gap of 32s. Such a slow time course is striking, for many context effects have been observed to decay much more rapidly. In the visual priming effect, it was observed that introducing a delay between the context and the test yielded to a rapid and exponential decrease of the bias. Pinkus & Pantle, consistently with previous results, reported that the decay of the priming effect occurs in a range of 500 ms to 1 s [10, 286], which correspond to refractory periods of neurons that are believed to support these assimilative context effects: once the baseline is recovered, the context effect fades out [152]. In the spectral-contrast effect, temporal decays of the context effect were reported to occur with hundreds of milliseconds [159, 232].

However, other studies have reported similarly slow decays. In motion perception, it was observed that the effect of an ambiguous motion does not decay after 5s of "blank" [172]. Similarly, Holt *et al.* observed that the effect of a long context was maintained even when a silent gap as long as 1,3s was introduced between context and target [158]. Unfortunately, in both cited studies, longer gap durations were not tested, thus not allowing the comparison with our results. The observed decay of this context effect is informative of the nature of the underlying neural processes: the persistence of the effect over seconds suggest that the locus of the underlying neural processes is not likely to be found in the primary sensory areas [234]. Indeed such remanence

can hardly be accounted for by recovery periods of peripheral neurons and seemingly point to processes stages known to display concordant recovery dynamics [386, 382].

Comparable temporal properties have been observed in studies on intermittent presentations of ambiguous stimuli [271]. Typically, an assimilative effect is observed between subsequent presentations of the same ambiguous stimulus, with a decay of the effect occurring over the course of minutes [46, 234]. This phenomenon is thought to reflect perceptual "stabilization" or "sensitization", a process that "freezes" the perception of an ambiguous stimulus in one of the possible interpretation for several minutes [222, 234, 236]. Snyder et al. observed a similar effect in the perceptual organization of ambiguous tone sequences in the streaming paradigm: the last perceived organization of tones tended to be maintained in the next presentation [352]. Interestingly, those effects were shown more robust over time than other context effects, lasting for seconds to minutes [222]. Given the extremely robust nature of the context effect presented in this chapter, it seems possible that similar processes are at play: the context effect may result from high-order mechanisms devoted to perceptual stabilization over time [234], a process likely to involve perceptual memory processes [277].

Although storage of the auditory information is arguably crucial in the observed effect, whether the remanence of this effect rests on entirely automatic sensory memory [72] or benefit from active and conscious recall from listeners remains unknown. Individual results indicate that listeners' abilities to maintain the context effect for the longest gap duration tested (64 s) vary. Possibly, this differences arise from distinct strategies during the silent gap, some listeners might be repeating over the context sequence in their head, while other might think of something completely different. To explore that question, further experiments exploring the effect of distracting tasks during the context and during the silent gap would be needed. Kanai & Vestraten conducted a study with similar questions and found that a distractive task during the encoding (context presentation) or the retention (silent gap) modulated the strength of the perceptual stabilization effect [173].

Exp. 1.3 was conducted to explore the interactions between the duration of context presentation and the duration of silent gap before test. We observed that the context effect was assimilative for all tested combinations of gap and context duration conditions. This experiment revealed an interaction between the two temporal factors that results in a rapid build up of the effect as the gap duration increases for short context durations. Namely, when the context duration is very short (20 and 40 ms), the context effect is strengthened by an increasing silent gap duration (0-250 ms). Similar findings were found in studies exploring the interaction of context and gap durations for the strength of context effects. However, as pointed out by Noest et al., such interaction effects may arise from the co-existence of contrastive and assimilative effects, the effect of the gap having different impact on facilitative and suppressive processes [268]. Our findings suggest that the interaction does not arise from distinct effects on competing context bias, since all results point to the existence of a single, assimilative, context effect.

Alternatively, the interaction resulting in the build up of context strength for increasing gap durations only for short durations could indicate that the neural processes involved are not located in primary sensory areas. Rather the fact that for the same amount of acoustic information the silent duration increases the bias induced on the ambiguous stimulus suggests that the neural processes are located at a high stage of the pathway, thus requiring more time to be fully operant [158]. This interpretation of the build up interaction is also consistent with the robust remanence over time which points to the involvement of high-level neural adaptations [386, 382].

To conclude, the present series of studies uncovered the outstanding temporal properties of a context effect on the perception of an ambiguous interval. It was indeed established by very brief tones and could remain effective even after a second of silence. Auditory objects are temporally-defined by definition and carry critical information over broad-range temporal scales, from milliseconds to minutes. The context effect under scrutiny in this chapter could contribute to highlight auditory processing requiring the integration of

information over long temporal scales, such as speech [183] and music [282] perception and is therefore a well-fitted tool to unravel the processes at play in everyday auditory perception.

Chapter 5

Ambiguity is perceived differently by musicians and non-musicians

The work described in this Chapter is published in: C. Pelofi, V. De Gardelle, P. Egré, and D. Pressnitzer. Inter-individual variability in auditory scene analysis revealed by confidence judgments (in press). *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Abstract

Because musicians are trained to hear out sounds within complex acoustic scenes, such as an orchestra playing, it has been hypothesized that musicianship improves general auditory scene analysis abilities. Here, we compared musicians and non-musicians in a behavioral paradigm using ambiguous stimuli, combining performance, reaction times, and confidence measures. We used “Shepard tones”, for which listeners may report either an upward or a downward pitch shift for the same ambiguous tone pair. Musicians and

non-musicians performed similarly on the pitch-shift direction task. In particular, both groups were at chance for the ambiguous case. However, groups differed in their reaction times and judgments of confidence. Musicians responded to the ambiguous case with long reaction times and low confidence, whereas non-musicians responded with fast reaction times and maximal confidence. In a subsequent experiment, non-musicians displayed reduced confidence for the ambiguous case when pure-tone components of the Shepard complex were made easier to hear out. The results suggest an effect of musical training on scene analysis: we speculate that musicians were more likely to hear out components within complex auditory scenes, perhaps because of enhanced attentional resolution, and thus discovered the ambiguity. For untrained listeners, stimulus ambiguity was not available to perceptual awareness.

Introduction

Two observers of the same stimulus may sometimes report drastically different perceptual experiences [42, 90]. When such inter-individual differences are stable over time, they provide a powerful tool to uncover the neural bases of perception [342, 171]. Here, we investigated inter-individual differences for auditory scene analysis, the fundamental ability to focus on target sounds amidst background sounds. We used an ambiguous stimulus [342], as inconclusive sensory evidence should enhance the contribution of idiosyncratic processes [187]. We further compared listeners with varying degrees of formal musical training, as musicianship has been argued to impact generic auditory abilities [259]. Finally, we combined standard performance measures with introspective judgments of confidence [80]. This method addressed a basic but unresolved question about ambiguous stimuli: are observers aware of the physical ambiguity, or not [371, 334]?

Ambiguous stimuli have been used to uncover robust inter-individual differences in perception. For vision, color [42] or motion direction [400] can be

modulated by strong and unexplained idiosyncratic biases, surprisingly stable over time. For audition, reports of pitch-shift direction between ambiguous sounds have also revealed stable biases [90], correlated to the language experience of listeners [92] although this has been debated [308]. Stable inter-individual differences also extend to so-called metacognitive abilities, such as the introspective judgment of accuracy of our percepts [119, 80]. When they are not treated as experimental noise, inter-individual differences can be leveraged to correlate behavioral characteristics subject-by-subject [388], and to probe neurophysiological (Kondo et al., in press), neuroanatomical [171], or genetic [174] bases of perceptual processing.

For auditory perception, one long-recognized source of inter-individual variability is musical training [409]. Musical training provides established benefits for music-related tasks, such as fine-grained pitch discrimination [259, 58]. Generalization to basic auditory processes is still under scrutiny, however. In particular, a number of studies have investigated whether musicianship improved auditory scene analysis. Musicians were initially shown to have improved intelligibility for speech in noise, which was correlated to enhanced neural encoding of pitch [274]. However, attempts to replicate and generalize these findings have been unsuccessful [317]. An advantage for musicians was subsequently found by emphasizing non-auditory aspects of the task, by using intelligible speech as masker instead of noise [368, 16], but null findings also exist with intelligible speech as masker [39]. For auditory scene analysis tasks not involving speech, musicians were better at extracting a melody [20] or a repeated tone [17, 272] from an interfering background. Musicians were also more likely to hear out a mistuned partial within a complex tone [411] or within an inharmonic chord [115].

Here, we take the comparison between musicians and non-musicians to a different setting, using ambiguous stimuli. We used Shepard tones [342], which are chords of many simultaneous pure tones, all with an octave relationship to each other. When two Shepard tones are played in rapid succession, listeners report a subjective pitch-shift, usually corresponding to the smallest log-frequency distance between successive component tones. When two suc-

cessive Shepard tones are separated by a frequency distance of half-an-octave, however, an essential ambiguity occurs: there is no shortest log-frequency distance to favor either up or down pitch-shifts. In this case, listeners tend to report either one or the other pitch-shift direction, with equal probability on average across trials and listeners [342]. Inter-individual differences have been observed for the direction of reported shift, but no systematic effect of musicianship [90, 92].

We did not investigate inter-individual differences in pitch-shift direction bias, but rather, differences in the introspective experience of Shepard tones: are listeners aware of the ambiguity, or not? The experimental measure of ambiguity so far has been an equal split between “up” and “down” pitch-shifts reports, for the same stimulus. However, there are three possible reasons for such an outcome. First, listeners may hear neither an upward nor a downward shift, and respond at chance. Second, listeners may hear simultaneously upward and downward shifts, and randomly choose between the two. Third, listeners may clearly hear one direction of shift, and report it unhesitantly, but this direction may change over trials. The question bears upon current debates on the nature of perception with under-determined information, which is the general case. The first two options, hearing neither or both pitch shifts, would be compatible with what has been termed vagueness: response categories are fuzzy, non-exclusive, and observers are aware of their uncertainty when selecting a response [299, 105]. The third option would be more akin to what is assumed for bistable stimuli [334]. We will designate this third option as a “polar” percept: observers are sure of what they perceive, unaware of the alternatives, but they differ in which percept they are attracted to. Based on informal observations, Shepard described ambiguous tone pairs as polar [342]. Deutsch further argued that the stable individual biases observed for such sounds pointed to polar percepts [90]. Interestingly however, the idiosyncratic biases can be overcome by hysteresis [63] or cross-modal influence [310], suggesting that both percepts may be available to the listener. Here we tested directly for the participants’ introspective confidence in their perception of Shepard tone pairs.

In three behavioral experiments, we investigated the perception of pitch-shifts between Shepard tone pairs, comparing ambiguous and non-ambiguous cases. We collected pitch-shift direction choices, but also reaction times and judgments of confidence. We hypothesized that, if listeners were aware of the perceptual ambiguity in physically ambiguous stimuli, this would translate into longer reaction times [371, 168] and lower confidence judgments. We finally controlled whether non-musicians could report the ambiguity when the component tones were made easier to hear out, through acoustic manipulation [85]. Results showed that the Shepard tones were polar for the naïve non-musicians listeners, but that musicianship or acoustic manipulation could reveal the ambiguity.

5.1 Experiment 2.1

In this experiment, a sequence of two tones, T1 and T2 were presented to listeners. The task was to report the pitch-shift direction of the interval. The interval T1-T2 was varied. In the 0st. interval condition, the sequence was composed of two identical tones so that no pitch-shift separated them. Hence, in this condition, we expected listeners to be at chance level in the direction response. It was previously reported that in the case of a 6st. interval, listeners also responded at chance, as they were as likely to report an upward or a downward pitch-shift while the response followed a proximity cue principle for other intervals [342]. However, the direction response does not disentangle two alternative hypothesis regarding the perceptual processing underlying the chance level response observed in 6st. intervals; it might either reflect a switch of the direction perceived among listeners from trials to trials, or it could result from uncertainty as to which direction was heard. As those two hypothesis would lead to observe the same response direction patterns, confidence ratings and response times were collected in order to tackle this yet unsolved question. The experimental design was:

Subj<Group2>*Interval12

5.1.1 Method

Participants

Sixteen self-reported normal hearing participants (age in years $M = 26.4$, $SD = 5.11$) were included in the experiment. Eight were musicians (four men and four women, age $M = 27.1$, $SD = 6.79$) and eight were non-musicians (four men and four women, age $M = 26.1$, $SD = 2.29$). Musicians had more than five years of musical practice in an academic institution. Among them, three were professionals (a clarinet player, a pianist, and a composer and pianist, all with self-reported absolute pitch). Among the non-musicians, four reported having no musical training whatsoever, the other four having practiced an instrument for less than four years in a non-academic institution. Such a binary distinction between musicians and non-musicians was arbitrary and based on the sample of participants who volunteered for the experiment. A finer sampling will be presented later with the online experiment. The two groups did not differ with respect to age (two-samples t -test, $t(14) = 0.39$, $p = 0.69$). All were selected after the second version of the [Screening procedure](#) and paid for their contribution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described in the previous chapter ([Stimuli](#)). Briefly, nine pure tones all with an octave relationship to a base frequency, F_b , were added together to cover the audible range. They were amplitude-weighted by a fixed spectral envelope, as illustrated in Figure 5.1.1 (Gaussian in log-frequency and linear amplitude, with a $M = 960$ Hz and $SD = 1 \log_2$ unit). The relative amplitudes of the components, $A(f)$, were computed using the equation below where f is the frequency of the component, c_f , the central frequency of the envelope and σ , the standard deviation.

$$A(f) = \exp\left(\frac{1}{2} * \left(\frac{\log(f/cf)}{\log(2)/\sigma}\right)^2\right)$$

A schematic representation of a Shepard tone spectrogram is displayed in Figure 5.1.1.

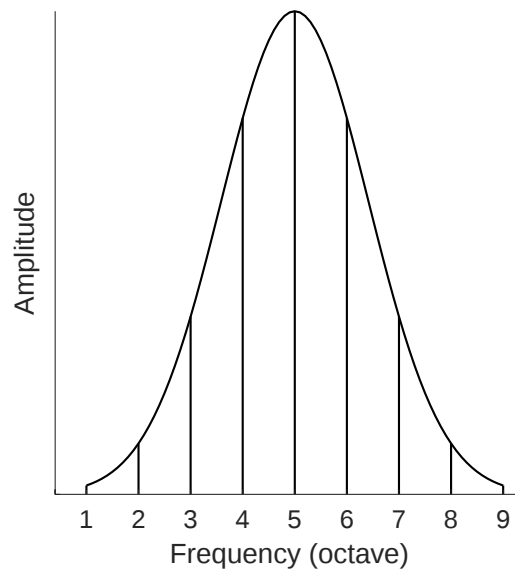


Figure 5.1.1: Shepard tones were generated by adding octave-related pure tones, weighted by a bell-shaped amplitude envelope.

A trial consisted of two successive tones, T1 and T2. The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz, to counterbalance possible idiosyncratic biases in pitch-shift direction preference [90, 63]. The Fb interval between T1 and T2 was randomly drawn, uniformly, from 0 semitones (st) to 11 st in steps of 1 st. The interval of 6 st corresponds to a half-octave, the ambiguous case. Other intervals were less ambiguous, as it was expected that listeners would be biased towards the pitch-shift direction of the shortest path. The duration of T1 and T2 was 125ms each, including 5ms raised-cosine onset and offset ramp. There was no delay between T1 and T2 with no delay between the two tones.

Each interval condition was presented forty times with presentation order

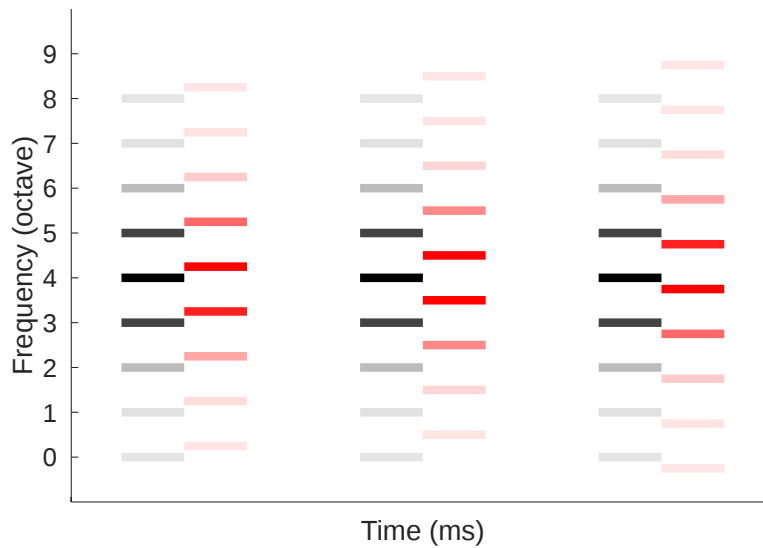


Figure 5.1.2: Pairs of Shepard tones, T1 and T2 for three interval conditions, 3 st, 6 st and 9 st.

randomly shuffled. As in previous experiment, to minimize across-trial effects participants were played an inter-trial sequence of five tones between trials. To minimize context effects [63], the inter-trial tones were designed as described in the previous chapter (*Stimuli*). The Fb for inter-trial tones was randomly drawn, uniformly between 60 Hz and 120 Hz. The experiment lasted for about 90 minutes, in a single session split over 4 sessions. Figure 5.1.2 displays a schematic representation of three trials with different interval conditions: 3 st, 6 st and 9 st.

Apparatus and procedure

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustics Company). Stimuli were played diotically through an RME Fireface 800 soundcard, at a 16-bit resolution and a 44.1 kHz sampling rate. They were presented through Sennheiser HD 250 Linear II headphones. The presentation level was 65 dB SPL, A-weighted. Participants

provided “up” or “down” response through a custom-made response box, which recorded reaction times with sub-millisecond accuracy. The right/left attribution of responses was counterbalanced across subjects. Trial presentation was self-paced. A trial was initiated by participants depressing both response buttons. This started a random silent interval of [50-850] ms followed by the stimulus pair T1 - T2. Participants were instructed to release as fast as possible the response button corresponding the pitch-shift direction they wished to report. Then, participants used a computer keyboard to rate their confidence in the pitch-shift direction report. They used a scale from 1 (very unsure) to 7 (very sure). The inter-trial sequence was then played and the next trial was ready to be initiated by listeners pressing buttons.

Data analysis for perceptual performance

For each listener and interval condition, the proportion of “up” responses $P(\text{up})$ was computed from the pitch-shift direction responses. Psychometric curves were fitted to individual listeners over the 1st-11th interval range, using cumulative Gaussians and estimating their parameters using the `psignifit` software [402]. The fitting procedure returned the point of subjective equality corresponding to $P(\text{Up}) = 0.5$; a noise parameter corresponding to the standard deviation of the cumulative Gaussian, σ , inversely related to the slope of the psychometric function; and the two higher and lower asymptotes. The response times (RTs) were defined relative to the onset of T2, the first opportunity to provide a meaningful response. RTs faster than 100 ms were discarded as anticipations. Because of the long-tailed distribution typical of RTs, the natural logarithms of RTs were used for all analyses.

Data analysis for metacognitive performance

We used an extension of the signal detection theory framework to quantify the use of the confidence scale by each participant [237]. In this framework, a stimulus elicits a value of an internal variable, and this value is

used to predict both perceptual decisions and metacognitive judgments of confidence. For perceptual decisions, as is standard with signal detection theory, the internal variable is compared to a fixed criterion value. For confidence judgments, it is the distance between the internal variable and the criterion that is used: values closer to the criterion should correspond to lower confidence. Assuming no loss of information, the metacognitive performance of each participant can then be mathematically derived from the perceptual performance. As shown in [237], this performance can be expressed as meta- d' , which measures the perceptual information (in d' units) that is translated into the empirical metacognitive judgments. Under those assumptions, meta- d' equals d' for participants with perfect metacognition. A complete formulation of the method is available in [237, 238] and comparisons with other techniques reviewed in [118]. For each participant, we computed d' , meta- d' , and the ratio meta- d'/d' which quantifies the efficacy of metacognitive judgments [118]. Intervals from 1 st to 5 st, which had an expected correct response of “up”, were treated as signal trials. Intervals from 7 st to 11 st, which had an expected correct response of “down”, were treated as noise trials. Intervals of 0 st and 6 st, for which there was no expected correct or incorrect response, were discarded from this analysis. For the confidence judgments, we summarized confidence levels using a median-split for each participant and added 1/4 trial to each condition (stimulus \times response \times confidence) to ensure that there were no empty cells in the analysis.

5.1.2 Results

Pitch-shift direction response

The results are displayed in Figure 5.1.3: the $P(\text{up})$ averaged across participants is plotted as a function of interval $T1 - T2$. As expected, both musicians and non-musicians reported mostly “up” for small intervals and “down” for large intervals, corresponding to the shortest log-frequency dis-

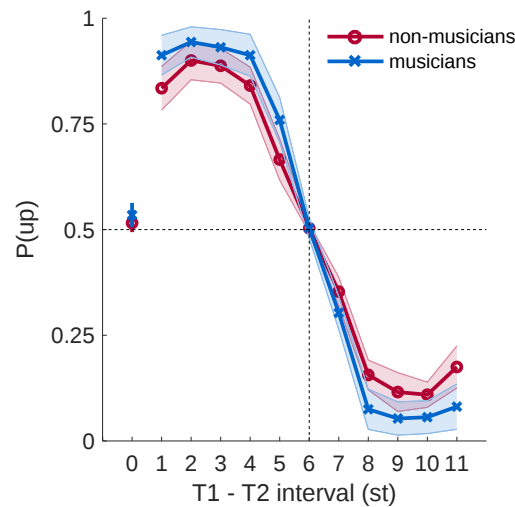


Figure 5.1.3: The proportion of “up” responses, $P(\text{up})$, is displayed for each interval and averaged within groups (non-musicians: red o, musicians: blue x). For all panels, shaded areas indicate ± 1 standard error about the mean. The interval of 6 st corresponds to the ambiguous case.

tances between successive components. Figure 5.1.4 and 5.1.5 displays $P(\text{up})$ and fitting curves for each listener in both groups. Performance was around the chance level of $P(\text{Up}) = 0.5$ for both the 0-st interval, corresponding to no physical difference between T1 and T2, and for the 6 st interval, corresponding to the ambiguous case with no shortest log-frequency distance.

Overall accuracy, as measured by σ , was statistically not different between musicians and non-musicians (two-sample t -test: $t(14) = 0.46$, $p = 0.16$). The trend for a higher performance for musicians at the extreme of the raw values (as visible in Figure XX) was not confirmed in the fitted psychometric functions (one-sample t -tests, upper asymptote: $t(14) = 1.46$, $p = 0.16$, lower asymptote: $t(14) = 1.28$, $p = 0.22$). The PSE corresponding to $P(\text{up}) = 0.5$ was not different from 6 st for both groups (one-sample t -tests, musicians: $M = 5.96$, $t(7) = -0.26$, $p = 0.80$; non-musicians: $M = 6.05$, $t(7) = 0.37$, $p = 0.72$) and not different across groups (two-sample t -test, $t(14) = 0.45$, $p = 0.66$).

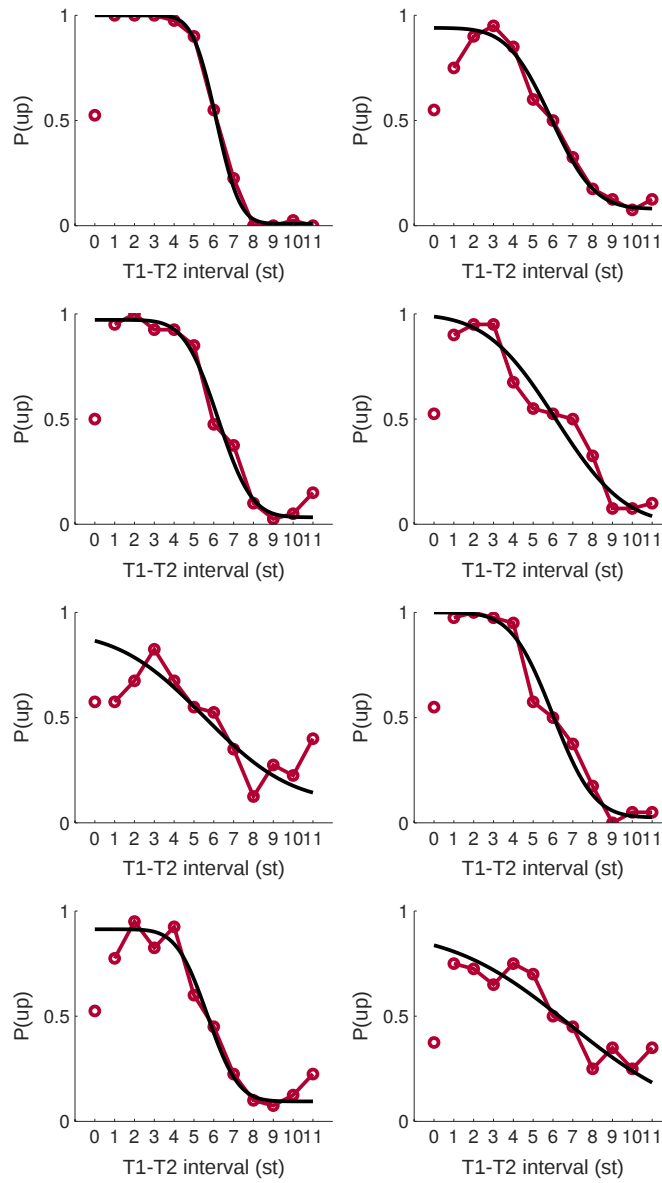


Figure 5.1.4: Individual results for non-musicians: in red, the proportion of “up” responses, $P(\text{up})$ is displayed for each interval and in black, the fitted psychometric curve.

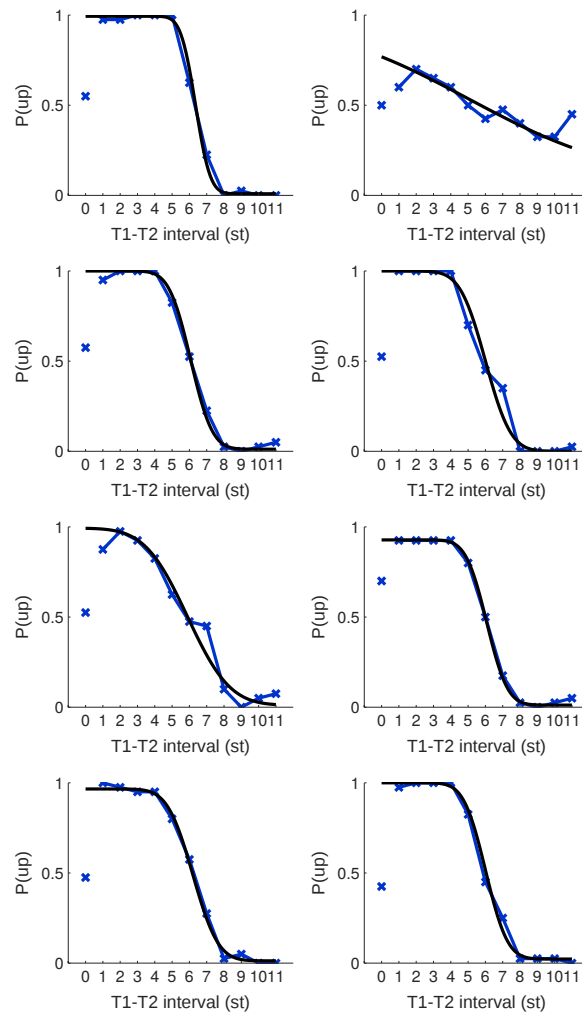


Figure 5.1.5: Individual results for musicians: in blue, the proportion of “up” responses, $P(\text{up})$ is displayed for each interval and in black, the fitted psychometric curve.

Response times

Figure 5.1.6 displays the averaged RTs as a function of the interval T1-T2. Perhaps predictably, the longest RTs for judging the pitch-shift direction were observed for 0 st, corresponding to identical sounds for T1 and T2. But for the other intervals, the pattern of response differed across groups. Musicians were slower for the ambiguous 6 st interval compared to less ambiguous intervals (1-5 st and 7-11 st), whereas non-musicians seemed faster for the ambiguous interval and slower for the non-ambiguous ones. Note that this ambiguous interval also contained the largest log-frequency distance between successive components of T1 and T2. We tested for the statistical reliability of this observation by averaging the log-RTs for non-ambiguous intervals (all intervals in the 1 st – 11 st range except 6 st) and comparing this value with the log-RT at 6 st. This difference between the averaged non-ambiguous cases and the ambiguous case, what we will now term the *ambiguity effect*.

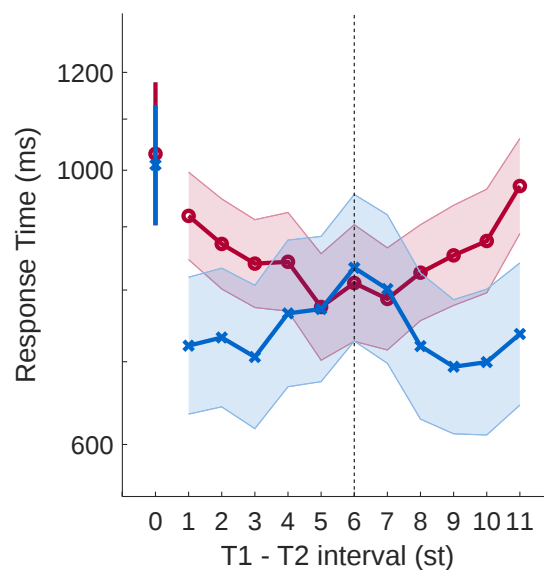


Figure 5.1.6: Response times for each interval, averaged within groups (non-musicians: red o, musicians: blue x). The natural logarithms of RTs were used to compute mean and standard error; y-labels have been converted to milliseconds for display purposes.

The results for the ambiguity effect for the two groups are displayed in Figure 5.1.7 and mean and SE are reported in Table 5.1. Comparing the ambiguity effect within each group confirmed that musicians were slower for the ambiguous interval compared to non-ambiguous intervals (paired t -test, $t(7) = 3.72$, $p = 0.007$). For non-musicians, this difference was not significant (paired t -test, $t(7) = 0.93$, $p = 0.38$). Finally, to test the interaction between response pattern and musicianship, we contrasted the ambiguity effect between the two groups. The contrast was significant (two-sample t -test, $t(14) = 3.07$, $p = 0.008$), confirming that musicians and non-musicians displayed different ambiguity effects in terms of log-RTs.

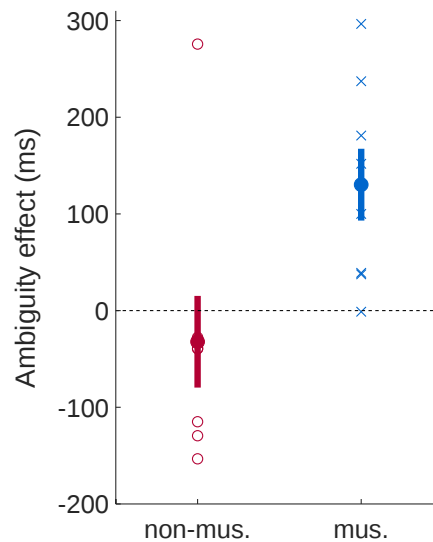


Figure 5.1.7: Ambiguity effect for log-RTs is obtained by averaging the log-RTs for non-ambiguous intervals (all intervals in the 1 st – 11 st range except 6 st) and comparing this value with the log-RT at 6 st within each group (non-musicians: red o, musicians: blue x). Bars indicate ± 1 standard error about the mean and individual symbols indicate individual results.

Chapter 5 Ambiguity is perceived differently by musicians and non-musicians

	Mean (SE)
non-musicians	-32 (134.1)
musicians	130.25 (105)

Table 5.1: Mean and standard error of ambiguity effect on Rt-logs for the two groups.

Confidence ratings

Did confidence ratings also indicate an ambiguity effect? First, averaged confidence ratings across groups were plotted as a function of the interval T1-T2, as displayed in Figure 5.1.8. Confidence was lowest for 0 st, and lower for musicians than non-musicians at this interval. For the other intervals, we observed different patterns of responses across groups. Musicians displayed a dip in confidence for the ambiguous case, but not non-musicians.

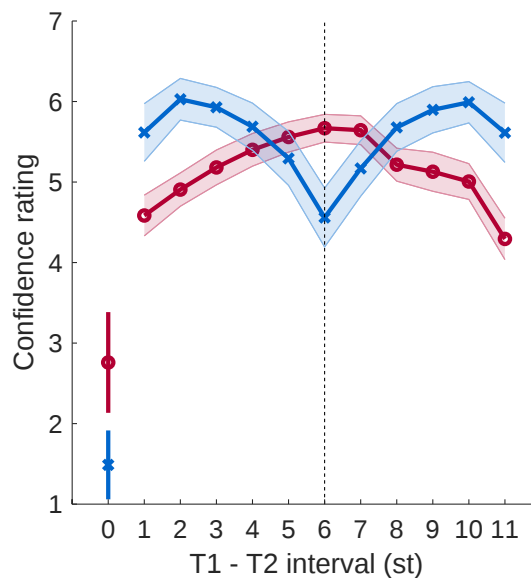


Figure 5.1.8: Confidence ratings on a scale to 1 (very unsure) to 7 (very sure) for each interval, averaged within groups (non-musicians: red o, musicians: blue x).

We used the same analysis method as above to quantify this observation: we averaged the confidence ratings for non-ambiguous intervals (all intervals

in the 1 st – 11 st range except 6 st) and compared this value with the confidence rating at 6 st. Results are displayed in Figure 5.1.9 and reported in Table 5.2. For musicians, confidence was higher for the non-ambiguous intervals compared to the ambiguous interval (one-sample t -test, $t(7) = 3.38$, $p = 0.012$). For non-musicians, the pattern was reversed, with higher confidence for the ambiguous interval (one-sample t -test, $t(7) = -3.34$, $p = 0.012$).

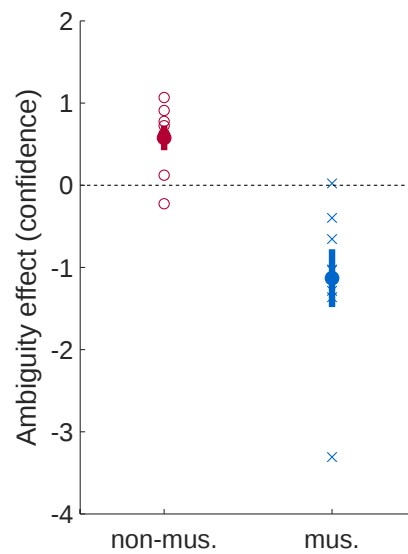


Figure 5.1.9: Ambiguity effect for confidence is obtained by averaging the confidence for non-ambiguous intervals (all intervals in the 1 st – 11 st range except 6 st) and comparing this value with the confidence at 6 st within each group (non-musicians: red o, musicians: blue x). Bars indicate ± 1 standard error about the mean and individual symbols indicate individual results.

	Mean (SE)
non-musicians	0.57 (0.42)
musicians	-1.08 (1.01)

Table 5.2: Mean and standard error of ambiguity effect on Confidence ratings for the two groups.

The interaction between group and ambiguity, tested with the ambiguity

effect, was significant (two-sample t -test, $t(14) = -4.43$, $p = 0.001$). Musicians and non-musicians displayed different ambiguity effects in terms of confidence.

Metacognitive performance

We investigated whether the difference in confidence judgments between non-musicians and musicians could be explained by a different use of the confidence scale. We estimated a meta- d' for each participant, which measures the perceptual information (in d' units) needed to explain the empirical metacognitive data, and expressed the results as the meta-ratio of meta- d'/d' [237].

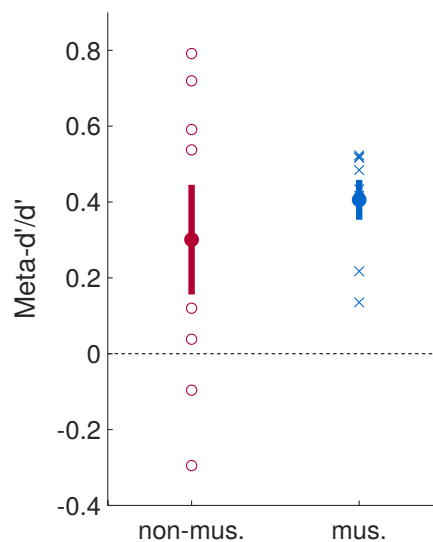


Figure 5.1.10: Meta- d'/d' for both groups (non-musicians: red o, musicians: blue x). Bars indicate ± 1 standard error about the mean and individual symbols indicate individual results.

We observed relatively low values of meta-ratio overall, and no difference between groups (musicians: $M = 0.4$, $SD = 0.14$; non-musicians: $M = 0.3$, $SD = 0.39$; $t(14) = -0.68$, $p = 0.5$). Moreover, there was no correlation between meta-ratio and ambiguity effect (Pearson correlation coefficient

$\rho(14) = 0.24, p = 0.36$). The metacognitive analysis thus suggests that not all of the perceptual evidence available for pitch-shift direction judgements was used for confidence judgements, but that, importantly, the efficacy of each individual participant in using the confidence scale was not related to the ambiguity effect. Results for this analysis are plotted in Figure 5.1.10.

Correlations of RTs and confidence

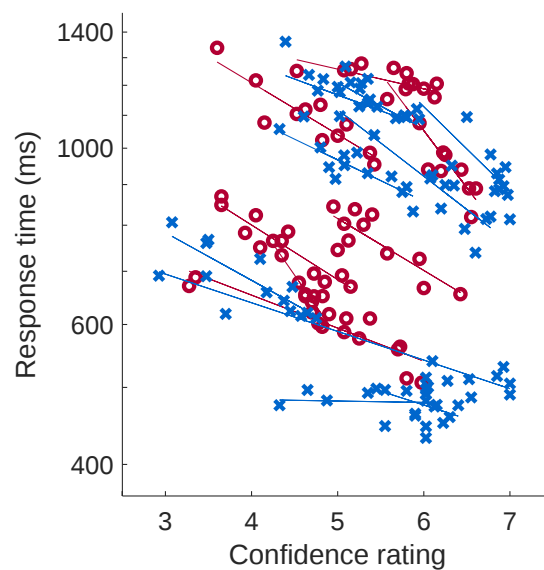


Figure 5.1.11: Correlation between confidence and log-RT. Each point represent an interval condition for an individual listener, solid lines are fitted linear regressions over all intervals for each individual listener (non-musicians: red o, musicians: blue x).

We hypothesized that RTs and confidence judgments would be negatively correlated. To test for this hypothesis, we correlated the log-RTs and confidence values for each participant separately, over the 1st-11th range. The correlations are displayed in Figure 5.1.11. We found negative correlations for almost all participants. The relation between the two variables was strong. Pearson correlation coefficient ρ averaged across participants: $M = -0.7$, $SD = 0.23$.

5.1.3 Summary

Exp. 2.1 was conducted in order to disentangle two distinct perceptual phenomena for the Shepard ambiguous interval: either an alternation of two unambiguous percepts on successive trials, or an uncertainty for each trial. Confidence ratings and response times were collected for varying degree of interval ambiguity in two different groups: musicians and non-musicians. Whilst performance in reporting the interval direction was not different for musicians and non-musicians, confidence ratings and response times strongly diverged between the two groups. Namely, musicians were more confident for intermediate intervals and less confident for the tritone, whereas non-musicians exhibited the opposite pattern. Response times corroborated this divergence. This strong difference between the two groups could not be explained by a different use of the confidence scale as assessed by the metacognitive analysis which revealed no difference between the groups.

5.2 Exp. 2.2: online experiment

In the main experiment, we observed different ambiguity effects for non-musicians and musicians. However, the comparison rested on a relatively small sample size (8 listeners in each group). Moreover, the binary definition of musicianship, required to analyze a small sample, could not reflect the spread of musical abilities between listeners. We performed an online experiment to try and replicate our main findings on a larger cohort. Participants completed a questionnaire about their music background. Secondly, they performed the pitch-shift direction task, followed by the confidence task, for a reduced set of intervals containing non-ambiguous and ambiguous cases. Response times were not collected, for technical reasons and also because they were strongly correlated with confidence in the main experiment. Thanks to the number of online participants with useable data ($N = 134$), we could then correlate the effect of ambiguity with years of musical training. Thirdly, a detection task of a "mistune" component within a

complex harmonic was conducted [411, 413]. Two versions of the mistune experiment were carried out. Here, only results from the second version will be presented. The aim was to investigate the correlation between ambiguity effect and ability to parse complex tones. Therefore, the online experiment consisted of three phases, the questionnaire, the Shepard tone experiment and the mistune experiment. In the Shepard tone experiment, the experimental design was:

Subj<Group2>*Interval6

In the mistune experiment, the experimental design was:

Subj<Group2>*Mistuned2

5.2.1 Method

Participants

Participants were recruited through an email call sent to a self-registration mailing list, provided by the “Relais d’information sur les Sciences Cognitives”, Paris, France. . They received an email explaining briefly the experiment and inviting them to complete it by following a link. The link led to the experiment platform where the entire procedure was explained to them (<http://cogitolabo.risc.cnrs.fr/pelofi/>). Participation was anonymous and listeners were not paid. In total, 359 participants started the experiment after reading the instructions (age: $M = 30.1$, $SD = 10.9$). Among them, 173 participants completed the Shepard tones experiment and 49 participants completed the second version of the mistune experiment. The high abandonment rate is explained by the duration the experiments, about 40-45 minutes.

Stimuli

Shepard tone experiment. All the tones used in this experiment were Shepard tones generated as described above (Stimuli). A trial consisted of two successive tones, T1 and T2. The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz. To keep the experiment at a reasonable length, the Fb interval between T1 and T2 were limited to [0, 2, 4, 6, 8 10] st. The interval of 6 st corresponds to a half-octave, the ambiguous case and the other intervals (2, 4, 8, 10 st) corresponded to less ambiguous cases. The duration of the test tones was 125ms each, including 5 ms raised-cosine onset and offset ramp. There was a delay of 125 ms between T1 and T2. Each interval was presented twelve times and the presentation order was shuffled across participants. Randomly interspersed with experimental trials were four catch trials, used as a screening procedure. Catch trials were constituted from two successive harmonic complex tones of 125 ms each (first 6 harmonics with flat spectral envelope) at an interval of 12 st. To minimize across-trial effects, an inter-trial sequence of three tones was played between trials. Inter-trials tones were generated as described above Stimuli. Each inter-trial tone had a duration of 125ms, including a 5ms raised-cosine onset and offset ramp.

Mistune experiment. The tones in this experiment consisted of harmonic complex that were created by adding together six pure tones of equal intensity and with integer-ratio relationship. Frequencies of the six pure tones components were 220, 440, 660, 880, 1100 and 1320 Hz. In half of the trials, the third component of the tone was mistuned by 2%, which corresponded to a frequency of 673.2 Hz. Harmonic complex had a duration of 150 ms, including a 10 ms onset and offset ramp.

Procedure

Participants were instructed to listen over headphones or loudspeaker, preferably in a quiet environment, but no attempt was made to control for the

sound presentation conditions.

Shepard tone experiment. After each trial, they provided a pitch- shift direction response by means of the up or down arrows on the keyboard. They were then asked for their confidence ratings on a scale ranging from 1 to 7 (1 = very unsure, 7 = very sure) using number keys on the keyboard.

Mistuned experiment. After each trial, they were asked to report if they heard one sound (control condition) or two simultaneous sounds (mistune condition). They reported their response by mean of the up or down arrows on the keyboard. They were then asked for their confidence ratings on a scale ranging from 1 to 7 (1 = very unsure, 7 = very sure) using number keys on the keyboard.

5.2.2 Results

Questionnaire

During the first phase, we collected information about age, education level, years of formal musical training in instrument and musical theory, current practice and type of instrument played. For simplicity, and to be consistent with the main experiment, we only used the duration of formal musical training in instrument as a shorthand for musicianship. In the self-selected sample of the Shepard tone experiment, mean number of years of musical training was $M = 8.74$, $SD = 8.49$. For group contrasts, we used the same criterion as for the main experiments: musicians were participants with 5 years or more of formal musical training. This resulted in a sample comprising 90 musicians and 83 non-musicians in the Shepard tone experiment and 23 non-musicians and 26 musicians in the mistune experiment.

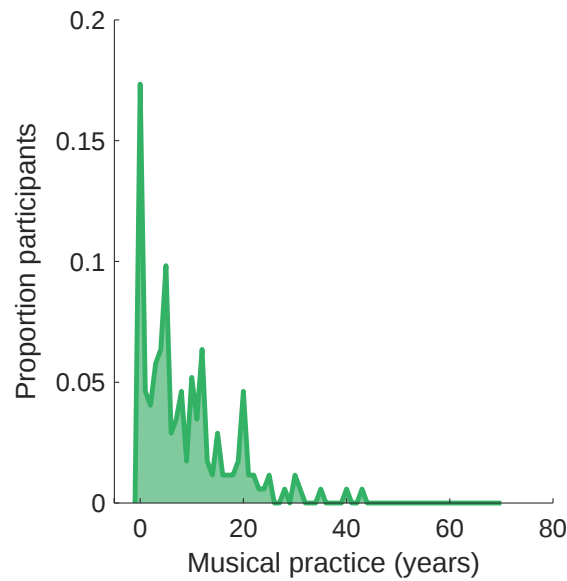


Figure 5.2.1: Histogram of the number of years of training in instrument from participants

Data-based screening

Data were expected to be noisy, especially as we did not control for the equipment used for playback nor for the attentiveness of listeners. Therefore, we chose to perform a data-based screening of participants, using their performance on the pitch-shift direction task (and not the confidence judgements). To this aim, psychometric functions were fitted for each participant, as described above. This fitting procedure returned three parameters: a point of subjective equality (PSE) corresponding to $P(\text{up}) = 0.5$, a noise parameter corresponding to the standard deviation of the cumulative Gaussian σ , inversely related to the slope of the psychometric function and the two higher and lower asymptotes. There was a large spread in the parameter representing accuracy σ ($M = 6.2$, $SD = 43.3$). After visual inspection of individual functions, we decided to discard participants with $\sigma > 5$. Indeed larger values of σ indicate shallower slope of the psychometric function and thus poorer performance. This selected 134 participants for subsequent analyses. The new sample comprised 86 musicians and 48 non-musicians. The unequal

balance may reflect higher performance or higher motivation from musicians in the online experiment. Performance on catch trials after the data-based screening was high overall, but marginally differed between groups as assessed by percent correct scores (musicians $M = 98.3$, $SD = 0.06$; non-musicians $M = 95.8$, $SD = 0.09$; two-sample t -test, $t(132) = -2.05$, $p = 0.04$).

Shepard tone experiment

Pitch-shift direction response

Figure 5.2.2 displays the proportion of "up" response as a function of the interval for the two groups, in red, non-musicians and blue, musicians. As in the previous experiment, participants responded in the expected way for unambiguous cases and were close to $P(\text{Up}) = 0.5$ for the ambiguous case.

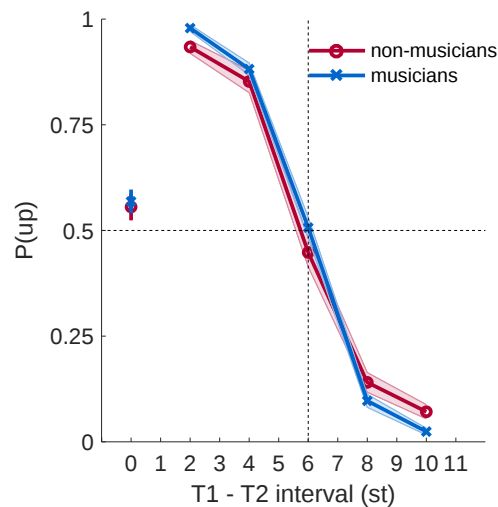


Figure 5.2.2: The thick lines represents the averaged proportion of "up" responses across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function each interval. The shaded area corresponds to the standard error.

Concerning overall accuracy as assessed by σ , no difference was found between the groups, as previously observed (two-sample t -test, $t(132) = 0.83$,

$p = 0.4$). However, as visible in the figure, groups differed for extreme interval values accuracy, corresponding to small log-frequency distances between T1 and T2, i.e. small intervals (two sample t -tests, upper asymptote: $t(132) = 2.72$, $p = 0.007$, lower asymptote: $t(132) = 3.60$, $p = 0.001$). This higher performance of musicians over non-musicians for accuracy in small pitch-shift is consistent with improved fine frequency discrimination for musicians [58].

Confidence ratings

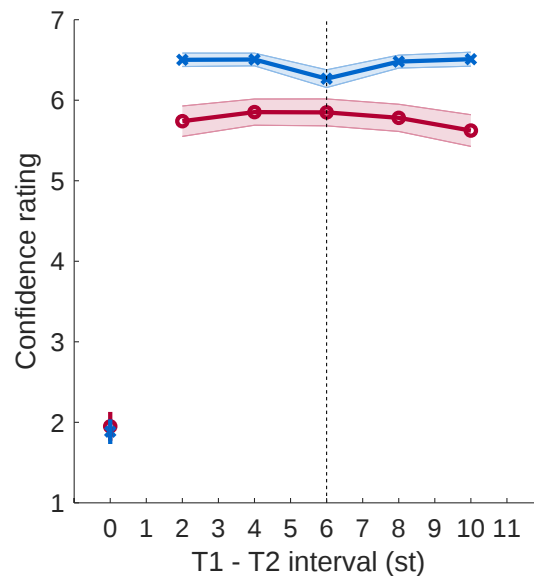


Figure 5.2.3: The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of each interval. The shaded area corresponds to the standard error.

Figure 5.2.3 displays averaged confidence rating as a function of the interval for both groups. Inspecting the figure, it appears that as in the previous experiment, musicians showed a dip in confidence for the ambiguous case, whereas non-musicians showed a broad peak in confidence for the same stimulus.

However, the contrast between intervals was less pronounced than in the main experiment. Moreover, musicians used higher confidence values than non-musicians overall. The same statistical analysis as for the main experiment was applied: the ambiguity effect was assessed by averaging confidence for non-ambiguous intervals of [2, 4, 8, 10] st and comparing it to confidence for the ambiguous interval of 6 st.

As visible in the Figure 5.2.4, musicians were less confident for the ambiguous interval (one-sample t -test, $t(85) = -4.22$, $p = 0.001$) and whereas non-musicians showed equivalent confidence for both types of intervals (one-sample t -test, $t(47) = 1.21$, $p = 0.23$). The two distributions indicate that the interaction between group and confidence was significant. Statistically, the contrast of ambiguity effect between the two groups confirmed this interaction (two-sample t -test, $t(132) = 3.45$, $p = 0.001$).

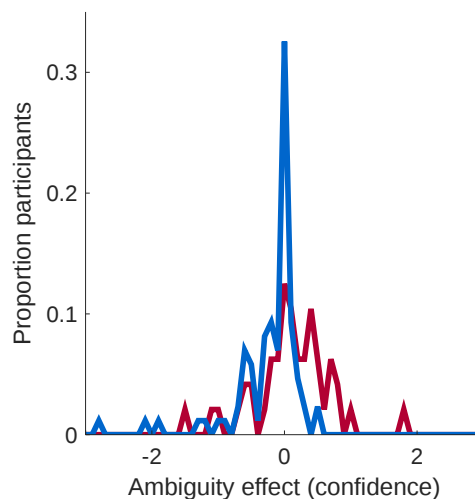


Figure 5.2.4: Histograms of the ambiguity effect, in confidence units, for non-musicians (red) and musicians (blue). The ambiguity effect was defined as the average confidence for all intervals excluding 0 st and 6 st, minus confidence for 6 st, the ambiguous case. Negative values correspond to listeners being less confident for the ambiguous case.

Thus, even though the differences across groups were less pronounced on average than in the main experiment, presumably because of the noisy nature

of online data, the main findings were replicated.

Metacognitive performances

The meta-ratio analysis revealed a difference between musicians and non-musicians (musicians: $M = 0.69$, $SD = 0.43$; non-musicians: $M = 0.36$, $SD = 0.62$, t -test, $t(132) = -3.58$, $p < 0.001$). However, the meta-ratio was not correlated to the ambiguity effect ($\rho(132) = -0.053$, $p = 0.55$). Also, we observed a sizeable proportion of participants with negative meta-ratio values, denoting a higher confidence for incorrect perceptual judgments, or meta-ratio values greater than 1, suggesting that confidence was not entirely based on perceptual information within a signal detection theory framework [237]. To test whether such unexpected response patterns affected the results, we performed an additional metacognitive analysis, retaining only those participants with a meta-ratio comprised between 0 and 1. In this sub-group of 52 musicians and 30 non-musicians, there was no difference in meta-ratio (musicians: $M = 0.57$, $SD = 0.26$; non-musicians: $M = 0.50$, $SD = 0.31$, $t(80) = -1.13$, $p = 0.26$). We also confirmed that the ambiguity effect was maintained for this sub-group ($t(80) = 2.67$, $p = 0.009$). As in the main experiment, there was therefore no link between the use of the confidence scale and the ambiguity effect.

Correlation with musical expertise

Given our large sample of participants exhibiting various degrees of musical expertise, we could analyze the relation between musical expertise and the ambiguity effect by observing the correlation between the two factors. The correlation is displayed in Figure 5.2.5. The rank-order correlation between the two variables was negative and significant ($\rho(132) = -0.25$, $p = 0.004$). Inevitably, the measure of "years of musical training" were partly confounded by the measure of "age". Also, one outlier subject with 25 years of musical training displayed an especially strong ambiguity effect, which could have

driven in part the correlation. We performed an additional correlation restricting the analysis to participants younger than 35 yrs, retaining 73 musicians and 41 non-musicians (and thus excluding the outlier, who was over 35 yrs). Age did not differ across the two subgroups (two-sample t -test, $t(112) = -0.97$, $p = 0.3$). Even when matching age, the correlation between years of musical practice and the ambiguity effect was maintained and became even more pronounced ($\rho(112) = -0.37$, $p = 0.0001$).

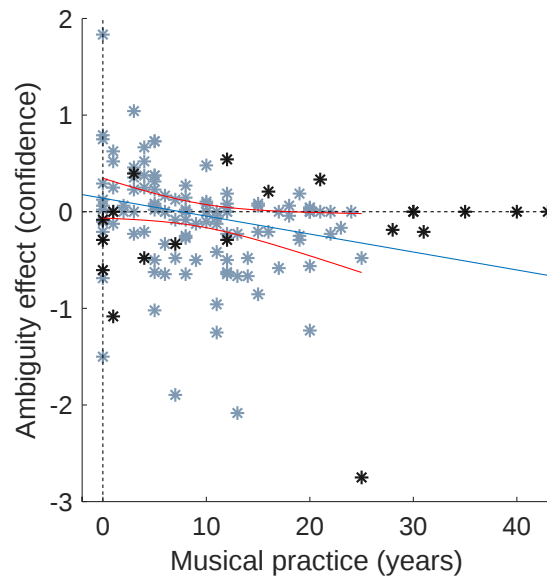


Figure 5.2.5: Correlation between the ambiguity effect and years of formal musical training. Blue stars indicate participants aged 35 yrs or less. Solid lines indicate the linear regression (blue line) and 95% confidence interval (red lines) fitted to the data for these younger participants.

Correlation with other factors

Finally, we looked at the relation between the ambiguity effect and all the other factors collected during the questionnaire for participants under 35. This analysis revealed that none of the other factor was more correlated to the tritone effect (i.e. the confidence difference measure) than the number of

years of musical training (plotted in Figure 5.2.6). The correlation (Spearman ranking) between the confidence difference and the number of weekly practice of a musical instrument was also significant ($\rho(132) = -0.19, p = 0.03$). None of the other factors had a significant link with the ambiguity effect (years of training in another musical discipline: $\rho(132) = -0.15, p = 0.07$, Academic level: $\rho(132) = -0.03, p = 0.71$).

Mistune experiment

Perceptual performance

d' for detection of the mistuned component was computed in a detection theory framework for each listener and averaged across groups. The averaged d' for both groups are displayed in Figure 5.2.6. Contrary to what was expected, musicians and non-musicians did not differ in terms of d' (musicians: $M = 2.33, SD = 1.74$, non-musicians: $M = 1.93, SD = 1.47$, two-sample t -test, $t(47) = -0.86, p = 0.39$).

Correlation with musical expertise

We analyzed the relationship between the sensitivity in the mistuning task, as assessed by the correlation between individual d' and ambiguity effect. The correlation is displayed in Figure 5.2.7. This analysis revealed that the ambiguity effect and the d' did not have any strong relationship: the rank correlation was not significant ($\rho(47) = 0.10, p = 0.47$).

5.2.3 Summary

Exp. 2.2. was conducted in order to probe the robustness of the tritone effect, as defined by the difference in confidence between the tritone interval and the intermediates intervals which diverge between the two groups.

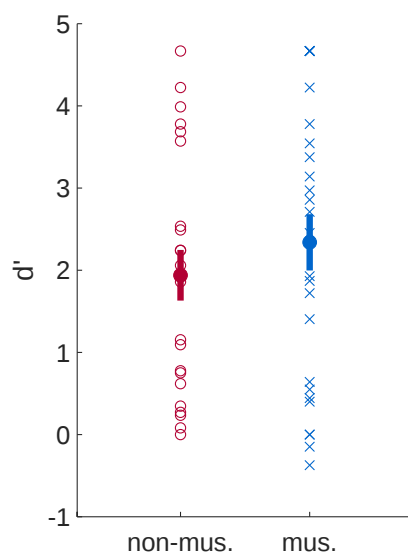


Figure 5.2.6: Each bar correspond to a the XX averaged across participants in both groups, in red, non-musicians and in blue, musicians. Error bars correspond to standard error.

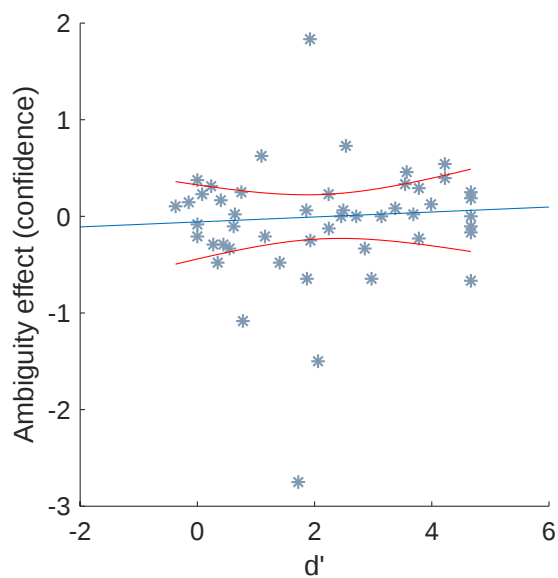


Figure 5.2.7: Correlation between the degree of musical expertise, measured as the number of years of instrument training, and the XX in the mistuning task.

Overall, the results strongly replicated those of Exp. 2.1., as the same group effect was observed. Besides, the large cohort of data yielded to observe the correlation between the degree of musical expertise (as defined by the number of years of an instrument practice) and the confidence difference between tritone and intermediate. We could establish that the more musical training participants had received, the less confident they were when reporting the direction of the ambiguous tritone, as compared to the intermediate intervals. Phase 3 tested the hypothesis that the “tritone effect” can be accounted by the greater abilities of musicians to segregate the concurrent components of the Shepard tones, leading them to perceive the pitch-shift direction ambiguity. But no difference in performance between the two groups.

5.3 Experiment 2.3

Exp. 2.1 and its online replication suggest that musicians were aware of the ambiguity at 6 st, whereas non-musicians were not. Following our initial hypothesis, this would be consistent with a perceptual difference between musicians and non-musicians: musicians were able to hear out the component tones of the ambiguous stimuli, which enabled them to discover the physical ambiguity of the stimulus. But it is also possible that the difference was in the report itself. Possibly, non-musician listeners also heard out the different component tones, but were unable to discover the ambiguity. We controlled for this possibility by making component tones easier to hear out for non-musicians, through acoustic manipulation. We compared conditions where all component tones started in synchrony, as in the previous two experiments, with conditions where the onset of each tone was jittered in time. Successive components of complex chords that would be normally fused can be explicitly compared when onset jitter is applied [85]. The hypothesis was that, if non-musicians could hear out component tones thanks to onset jitter, they would behave as the musicians of the main experiment. The experimental design was:

Subj<Group2>*Interval6*Jitter4

5.3.1 Method

Participants

Sixteen participants (age: $M = 24.25$, $SD = 0.79$), six men and ten women were included in this experiment after providing written informed consent. All had less than five years of formal musical training and were thus labelled non-musicians by our criterion. They were paid for their participation. Their hearing was tested through an audio-gram. All participants had thresholds below 20dB SPL for frequencies between 125Hz and 8kHz.

Stimuli

All the tones used in this experiment were Shepard tones generated as described above ([Stimuli](#)). Each trial consisted of the succession of two successive tones, T1 and T2. The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz, to counterbalance possible idiosyncratic biases in pitch-shift direction preference [90, 63]. As in the online experiment, the Fb interval between T1 and T2 were [0, 2, 4, 6, 8 10] st. The interval of 6 st corresponds to a half-octave, the ambiguous case and the other intervals (2, 4, 8, 10 st) corresponded to less ambiguous cases.

Three values of temporal jitter were prepared. For 0-ms jitter, stimuli were T1 - T2 test pairs generated as in the main experiment. For 50-ms jitter, a time jitter was introduced on the onset of individual component tones. Nine different onset-time values, linearly spaced between 0 ms and 50 ms included, were assigned, at random on each trial, to the components of T1 and T2 independently. For 100-ms jitter, the jitter values were linearly spaced between 0 ms and 100 ms included. In order to minimize any across trial effect, an inter-trial sequence of five inter-trial tones was played, as previously

described [Stimuli](#). Figure 5.3.1 displays a schematic representation of three trials.

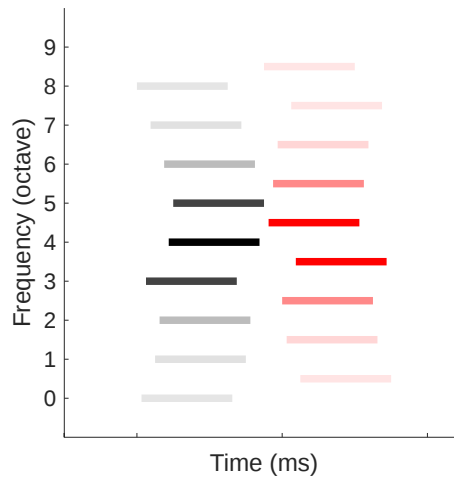


Figure 5.3.1: Schematic representation of stimuli in Exp. 2.1. Each frequency component in T1 and T2 is "jittered".

Apparatus and procedure

Listeners were tested individually in a sound-insulated booth. They performed six blocks, with a short rest after each block. The total duration of the experiment including the rests was approximately three hours, performed in a single session. During the first two blocks, only stimuli with 0-ms jitter were presented, for all intervals, with 40 repeats per interval. This was intended as a replication of the main experiment for this group of naïve listeners. We term these initial two blocks the "baseline" condition. For the following four blocks, all intervals and jitter values (0 ms, 50 ms, 100 ms) were presented at random, with 40 repeat per interval and jitter value. We designate these conditions by their jitter values. At the end of each trial, listeners were asked to report whether the interval formed by the two final tones was going upward or downward in pitch. To give their answer, they had to press either key 1 (downward shift) or key 2 (upward shift) of the

keyboard. The inter-trial sequence was then played. Response times were not recorded.

5.3.2 Results

Pitch-shift direction response

The proportion of "up" responses $P(\text{up})$ as a function of the interval condition and for the baseline and each jitter condition was computed and averaged across participants, as displayed in Figure 5.3.2. As in previous experiments, psychometric curves were fitted to $P(\text{Up})$ for each individual participant in all conditions, baseline and jitter conditions. The accuracy of participants, as measured by σ for the fitted psychometric functions, was equivalent for baseline and 100 ms jitter (pairwise t -test, $t(15) = -1.01$, $p = 0.327$). There was also no difference for the point of subjective equality (t -test, $t(15) = 0.40$, $p = 0.7$), which was not different from 6 st in both cases (baseline: $M = 5.97$, $SD = 0.28$, $t(15) = -0.46$, $p = 0.65$; 100 ms jitter: $M = 6.03$, $SD = 0.52$, $t(15) = 0.30$, $p = 0.76$). In summary, temporal jitter had no measurable effect on the pitch-shift direction task.

Confidence ratings

Confidence ratings are displayed as a function of T1-T2 interval and averaged across participants in Figure 5.3.3. Baseline results replicated the main and online experiments for non-musicians, whereas a dip in confidence appeared for those same non-musicians at a jitter of 100 ms. As before, we evaluated the ambiguity effect: we compared confidence for the non-ambiguous cases ([2 4 8 10] st) to confidence for the ambiguous case (6 st). For the baseline condition, we found no ambiguity effect ($t(15) = -0.10$, $p = 0.920$). To quantify the effect of jitter on a subject per subject basis, we normalized the ambiguity effect observed for each jitter conditions by subtracting the ambiguity effect observed for the baseline condition.

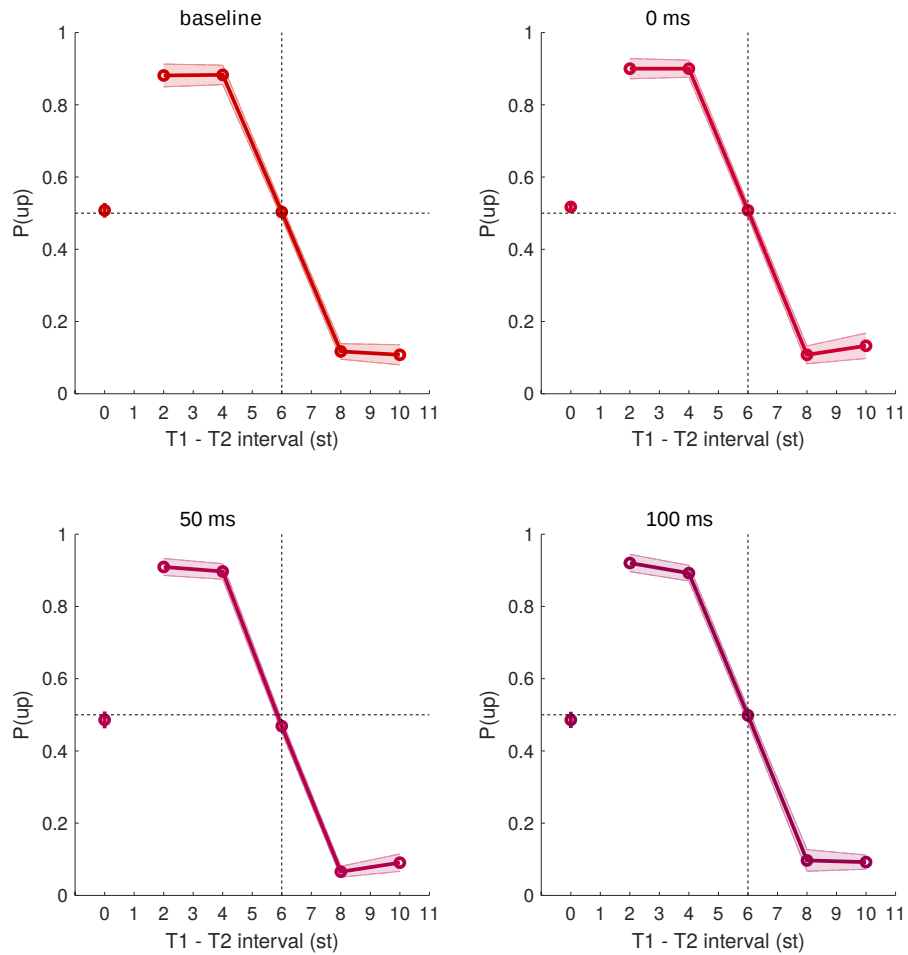


Figure 5.3.2: Averaged "up" response averaged across participants as a function of T1-T2 interval. The different plots correspond to the baseline and each jitter condition (error = SE).

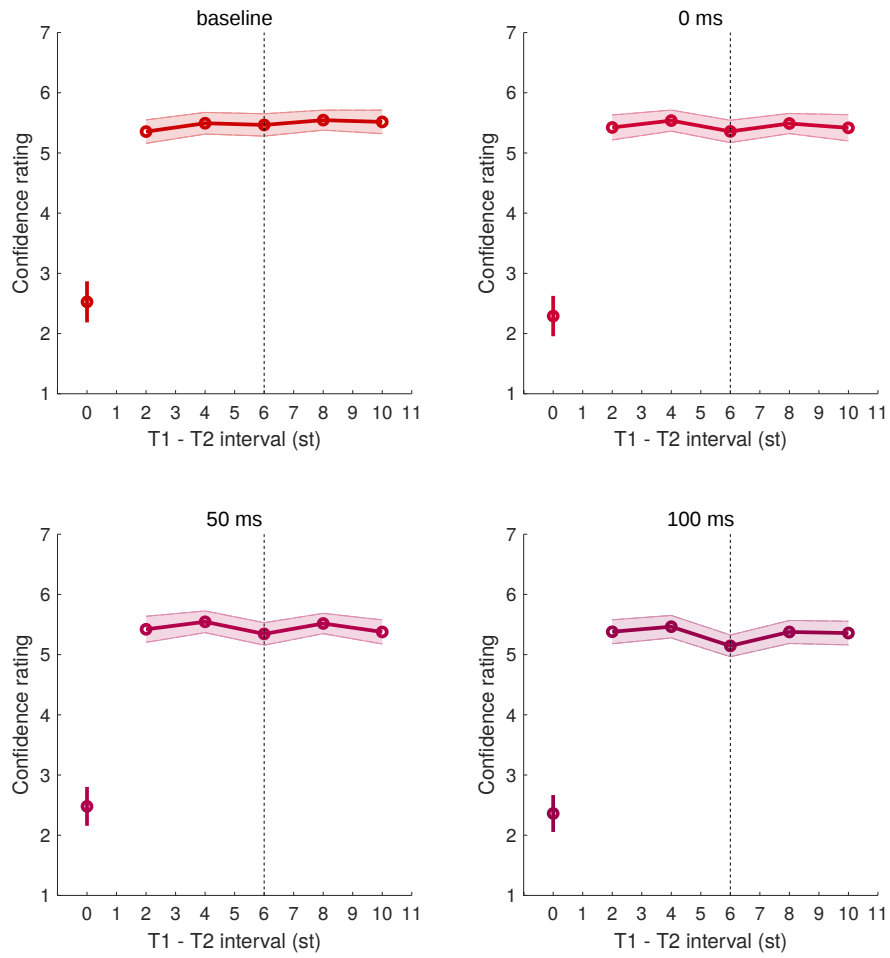


Figure 5.3.3: Averaged confidence ratings averaged across participants as a function of T1-T2 interval for the baseline and each jitter condition (error = SE).

Figure 5.3.4 displays the individual data for the normalized ambiguity: the ambiguity effect is normalized by the baseline for each participant. A negative ambiguity effect corresponds to a decrease of confidence for the ambiguous case and thus mimic the behaviors of musicians in the previous experiments. Not all participants exhibited such a negative ambiguity effect, but for some of them, the effect was as large as for musicians. When comparing the ambiguity effect to the baseline value, we only found a statistical difference for the 100 ms condition (0 ms: $t(15) = -1.40$, $p = 0.182$; 50 ms: $t(15) = -1.18$, $p = 0.258$; 100 ms: $t(15) = -2.52$, $p = 0.024$). However, when contrasting the three jitter conditions against each other, we did not find any significant difference (all pairwise comparisons $p > 0.17$).

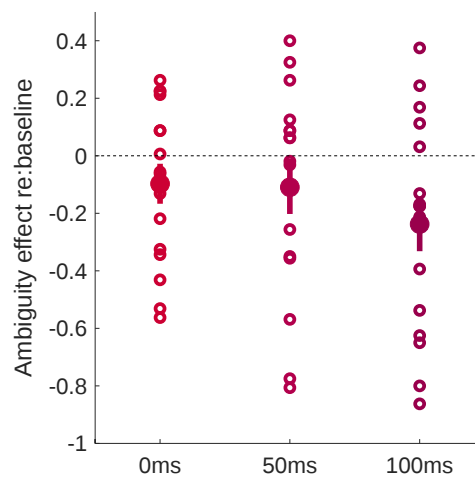


Figure 5.3.4: For each participant, we normalized the ambiguity effect observed in each condition by subtracting the ambiguity effect observed for the baseline. The panel displays individual means and standard errors about the means. Negative values signal a dip in confidence for the ambiguous case, when jitter was applied.

Metacognitive performances

There was only one group of participants in this experiment, but we still tested whether the ambiguity effect might have been related to a different use of the confidence scale across conditions. This was not the case, as variations in the ambiguity effect between baseline and 100-ms jitter were not correlated to variations in meta-ratio ($\rho(14) = 0.24$, $p = 0.38$).

5.3.3 Interim discussion

Non-musicians exhibited a significant confidence dip for the non-ambiguous interval when a large temporal jitter was introduced, while for the baseline condition no dip was observed. However, there was some variability across participants. Furthermore, results were statistically less clear cut when all jitter conditions were interleaved (note also that we did not correct for multiple comparisons). The lack of difference across jitter conditions is likely due to a trend towards an ambiguity effect for 0 ms, even though this condition corresponded the same stimulus as for the baseline. A possible interpretation of this contextual effect is that non-musicians were able to partially hear out the component tones, even without jitter, but only when their attention had been drawn to the stimulus' structure by other trials that contained jitter. A further point to consider is that we did not test whether the jitter values were sufficient for all participants to hear out component tones, which may also account for part of the variability across listeners.

5.4 Discussion

Results from the three experiments can be summarized as follows. In the main experiment, all participants responded with “up” and “down” responses in equal measure when judging the pitch direction of an ambiguous stimulus. The same pattern of chance result was, predictably, observed when

they compared two physically identical stimuli. However, confidence differed between the two cases: participants were more confident in their judgment for the ambiguous tones than for identical tones. Non-musician participants were even more confident for ambiguous tones, for which they were at chance, than for non-ambiguous tones, for which they could perform the task accurately. As the ambiguous tones also contained the largest interval between tones, results for non-musician participants are consistent with confidence mirroring the size of the perceived pitch shift, irrespective of stimulus ambiguity. Musicians provided the same pattern of responses for “up” and “down” judgments as non-musicians, but, importantly, response patterns differed markedly for confidence ratings. For musicians, confidence was lower for the ambiguous tones than for the non-ambiguous tones. In both groups of participants, response times mirrored the confidence ratings, with higher confidence corresponding to faster responses. The online experiment replicated those findings on a larger cohort ($N = 134$), further demonstrating that the ambiguity effect was correlated with years of musical training. The control experiment showed that at least some naïve non-musicians, initially unaware of the ambiguity, could exhibit an ambiguity effect when component tones were easier to hear out. This confirms that the difference between musicians and non-musicians was perceptual and not purely decisional.

What was the nature of the perceptual difference between musicians and non-musicians? Based on prior evidence [411, 115] combined with the results of Empt 2.3, we hypothesize that musicians were better able to hear out acoustic components within Shepard tones. Hearing out components would have revealed that two opposite pitch-shift directions were available, entailing low confidence in the forced-choice task. The advantage previously demonstrated for musicians when hearing out tones within a chord [115] or a complex tone background [272] did not seem to be based on enhanced peripheral frequency selectivity (although see [25]). In our case, the octave spacing of Shepard tones was also greater than the frequency separation required to hear out partials [254], so we can rule out a difference in the accuracy of peripheral representations.

Another possible difference between listeners, perhaps due to more central processes, is the distinction first introduced by Helmholtz between “analytic” versus “holistic” listeners [332]. Stimuli containing frequency-shifts changing in one direction if one focuses on individual component, or the opposite direction if one focuses on the (missing) fundamental frequency, have been used to characterize such a difference. Analytic listeners focus on individual components, whereas holistic listeners focus on the missing fundamental. In a first study, the analytic/holistic pattern of response was found to be correlated with brain anatomy, but not musicianship [331]. A later behavioral investigation suggested, in contrast, that musicians were overall more holistic [338]. Recent data indicate that listeners may in fact change their listening style according to the task or stimulus parameters [215]. In our case, musicians would be classified as analytic, as they spontaneously heard out component tones within ambiguous stimuli. Non-musicians were able to switch from holistic to analytic, when component tones were easier to hear out. This confirms that the holistic/analytic description is task-dependent [215].

We would argue that a more general characteristic is relevant to the inter-individual differences we observed: the ability to focus on sub-parts of a perceptual scene. In the visual modality, experts such as drawing artists [285] or video-game enthusiasts [138] have been shown to better resist visual crowding (the inability to distinguish items presented simultaneously in the visual periphery even though each item would be easily recognized on its own). This has been interpreted as enhanced “attentional resolution” for experts. The notion of attentional resolution can be transposed to auditory perception, through what has been termed “informational masking” [180, 102]. Informational masking is defined as the impairment in detecting a target caused by irrelevant background sounds, in the absence of any overlap between targets and maskers at a peripheral level of representation [180]. An effect of auditory expertise has been shown for informational masking. Musicians are less susceptible to informational masking than non-musicians, when using complex maskers such as random tone clouds with a

large degree of uncertainty [272], environmental sounds [58], or simultaneous speech [368]. This has been interpreted as better attentional resolution for musicians [272]. Importantly, informational masking can also occur in simpler situations, through a failure to perceptually segregate target and background [180]. The Shepard tones we used certainly challenged perceptual segregation, because component tones had synchronous onsets and were all spaced by octaves [47]. Scene analysis processes related to informational masking may thus have caused non-musicians to hear Shepard tones as perceptual units, while reduced informational masking may have helped musicians to hear out component tones. Moreover, in the control experiment, asynchronous onsets between components were introduced, a manipulation that has been shown to reduce informational masking [266]. This led non-musicians to behave qualitatively like musicians, again consistent with an implication of informational masking in our task.

We now come back to the distinction, drawn in the introduction, between vague and polar percepts. The results suggest that, for non-musicians, perception was polar: the intrinsic ambiguity of the stimuli was unavailable to awareness. This is consistent with recent observations using bistable visual stimuli [371]. In this visual study, the degree of ambiguity of bistable stimuli (defined as the proximity to an equal probability to report either percept) was varied, and reaction times were collected for the report of the first percept. No increase in reaction time was observed for more ambiguous stimuli, consistent with an unawareness of ambiguity, as in our results. Why would ambiguity, a potentially useful cue, not be registered by observers? One hypothesis is that ambiguity is not a valid property of perceptual organization at any given moment; either an acoustic feature belongs to one source, or it does not. This has been formulated as a principle of exclusive allocation for auditory scene analysis [47, 345]. However, there are exceptions to this principle, as for instance when a mistuned harmonic within a harmonic complex is heard out from the complex, but still contributes to the overall pitch of the complex [256, 254]. Another hypothesis, perhaps more specific to our experimental setup, involves perceptual binding over time. Shepard tones

are made up of several frequency components, which have to be paired over time to define a pitch shift. If musicians heard Shepard tones as perceptual units, this may have biased the binding processes towards a single and unambiguous direction of shift. In contrast, if musicians were able to hear out components, they may have experienced contradictory directions of pitch shift and thus ambiguity.

To conclude, we must point out that these interpretations remain speculative, especially as we have no direct evidence that musicians heard out component tones. Further tests could include counterparts to our control experiment, increasing informational masking to prevent musicians from hearing out component tones. This could be achieved for instance by presenting shorter duration sounds. Context effects could also be used to bias perceptual binding in a consistent manner across components [63], again predicting higher confidence for musicians. Irrespective of the underlying mechanisms, the mere existence of polar percepts demonstrates that a seemingly obvious feature of sensory information, ambiguity, is sometimes unavailable to perceptual awareness. The precise conditions required for polar perception remain to be explored experimentally. Ambiguity in the stimulus often causes multistable perception [334], but it is yet unknown whether all multistable stimuli elicit polar perception, or whether all polar percepts are associated to spontaneous perceptual alternations over time. By analogy to categorical perception, it could also be tested whether polar perception is a feature of conscious processing and absent from subliminal processing [79]. Finally, we showed that ambiguity was not experienced in the same way by different observers, so polar perception may be another useful trait to consider when investigating inter-individual differences.

In conclusion, ambiguous auditory stimuli [90, 342, 63] were judged with high confidence and fast reaction times by naïve non-musician listeners, showing that those listeners were unaware of the physical ambiguity of the sounds. This confirms an untested assumption in previous reports about those stimuli [90, 342, 63, 311]. In contrast, musicians judged ambiguous stimuli with less confidence and slower reaction times, suggesting that they did

perceive the ambiguity. We interpreted this inter-individual difference as an enhanced attentional resolution in crowded perceptual scenes for musicians [272, 411, 115]. From a methodological perspective, we showed robust effects in confidence judgements for matched performance within subjects (identical sounds versus ambiguous sounds), and robust inter-individual differences for matched performance and matched stimuli (musicians versus non-musicians). Ambiguous stimuli may thus provide a useful tool for probing the neural bases of inter-individual differences in perception.

Chapter 6

Context effect and perceived ambiguity: when musicians and non-musicians accord

The work described in this Chapter is reported in: C. Pelofi and D. Pressnitzer. Context-dependent variability in auditory scene analysis revealed by response times and confidence judgements (in preparation).

Abstract

When confronted to an ambiguous stimulus, whose inconclusive sensory evidence does not lead to a stable perceptual output, perceptual systems' sensitivity to contextual information is enhanced [190]. In the "Shepard illusion" [342], an ambiguous interval has been reported to be extremely sensitive to context information [63]. Here, we explored the effect of the context on the introspection of the exclusivity of the two alternative percepts. Listeners had to report the pitch-direction of an ambiguous interval preceded by a context that varied in strength. Confidence ratings and response times

were collected, as marker of the ambiguity perceived by listeners. Musicians and non-musicians were compared, as recent findings established a strong divergence between the two groups in terms of the perceived ambiguity (see preceding Chapter). Two experiments were conducted, in which the strength of the context was varied through temporal and spectral parameters, respectively. The results show an effect of context on the perceived ambiguity: as the direction was more biased towards one pitch-shift direction, due to the influence of the preceding contextual sequence, listeners from both groups report higher confidence judgment paralleled by faster response times.

Introduction

Perceptual ambiguity, where a conflicting set of information leads to unstable percepts, is a general perceptual phenomenon, observed in different sensory modalities [74, 61, 414]. Encompassing the wide range of findings, Leopold & Logothetis identified three fundamental hallmarks of perceptual ambiguity: randomness, inevitability and exclusivity [221]. Exclusivity corresponds to the fact that, while a stimulus leads to several alternative percepts, only one at a time is experienced. This could be rooted in ecological reasons: while perceptual systems process highly complex and sometimes contradictory set of informations, its intertwined relationship with action may constrain the perceptual outputs into being unambiguous [153, 145].

Exploring the introspection paralleling the perception of ambiguous stimuli, it has been reported that the ambiguity was processed unbeknownst to listeners [371]. Contradictory findings have reported that observers may report mixture percepts associated with greater ambiguity [157, 188, 251, 348]. However, none of those studies explored the influence of context effects on the prevalence of mixture percepts, although they are known to greatly impact the perceptual output of ambiguous stimuli [334, 190].

In Chapter 4, the temporal dynamics of a context effect on the perception of an ambiguous auditory interval were explored. The findings revealed very

short establishment properties and very long lasting effect. This remarkable robustness suggests that the underlying processes occur beyond primary sensory areas and require a storage of the perceptual trace at a high level of processing [234]. Also, for all the tested conditions, the effect was assimilative: the reported pitch-shift encompasses the frequency region corresponding to context components. It was hypothesized that the prior context resulted in context-induced tracks that "forced" to assign the octave-related components of the Shepard tones within the same track, thus biasing the pitch-shift in the same direction for all listeners [2].

In Chapter 5, the perceived ambiguity resulting from the interval was explored, using a combination of response times and confidence ratings. A strong and robust effect of musicianship was found: musical expertise was associated with a the perception of the ambiguity. This effect was thought to reveal enhanced "attentional resolution" of the sub-parts of Shepard tones. Namely, musicians being less sensitive to informational masking [272, 58, 368] caused by the octave-related components of the Shepard tones, they are more likely to hear competing pitch-shift directions and thus, report an ambiguous percept.

In this chapter, we explored the effect of context bias strength on the perception of ambiguity associated with the ambiguous interval. We hypothesized that the context sequence forces the attribution of octave-related component to the same "context-induced track" for musicians and non-musicians, resulting in a bias in the reported pitch-shift. However, the context bias should induce an additional effect for musicians: since the distinct octave-related components of the Shepard tones are attributed to the same context-induced tracks, the perceived ambiguity -thought to result from conflicting tracks of pitch-shifts- should decrease as the context bias increases.

To test this hypothesis, two experiments were conducted. In each, musicians and non-musicians were asked to report the pitch-shift of an ambiguous interval preceded by a context sequence whose strength was varied by temporal and spectral parameters. Confidence ratings and response times were

collected to assess the degree of perceived ambiguity associated with the ambiguous stimulus. In Exp. 3.1, the strength of the context was manipulated by varying the duration of one context tone (see Chapter 4). In Exp. 3.2, the strength of the context was manipulated by varying the frequency of one context tone.

6.1 Experiment 3.1

Expmt 3.1 was conducted in order to investigate the effect of context effect strength on the perceived ambiguity in the ambiguous interval. To this aim, we presented listeners with context and test trials in which the duration of the context was systematically varied in duration. At the end of the trial, listeners reported the pitch-shift direction of the final ambiguous interval, as rapidly and accurately as possible. Then, they were asked to report their confidence in their response using a confidence scale. The experimental design was:

Subj<Group2>*Context6

6.1.1 Method

Participants

Sixteen self-reported normal-hearing listeners ($M = 24.9$, $SD = 4.28$), nine men and seven women, participated in the experiment. One participants had not previously taken part in experiments involving Shepard tones and three only participated to Exp 3.2 (which was conducted before this one and with most of the participants in common with Exp 3.1). All participants passed the screening test, version 2 (see [Screening procedure](#)). Eight were musicians and eight were non musicians. The selection criteria for inclusion in the musician group was as follows: musicians had more than five years of musical practice in an academic institution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described in previous chapters ([Stimuli](#)). A trial consisted in the succession of a context tone (C) and two test tones (T1 and T2). The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz, to counterbalance possible idiosyncratic biases in pitch-shift direction preference [90, 63].

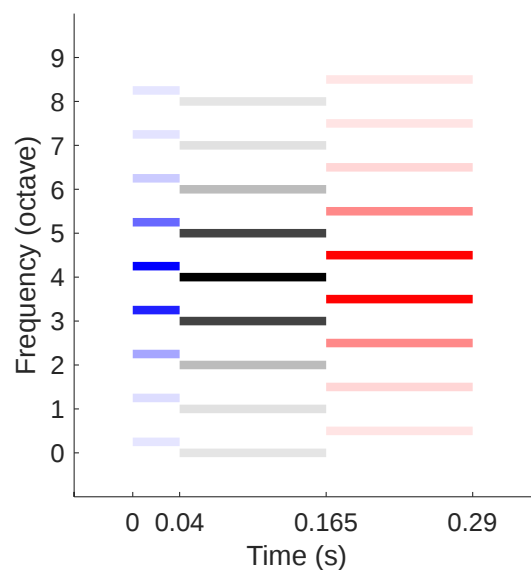


Figure 6.1.1: In a trial, a context tone preceded a pair of Shepard tones, T1 and T2 separated by a 6 st interval. When reporting the pitch-shift between T1 and T2, listeners tend to choose the path that encompasses the region where the context tone is. Here it would result in an upward shift response. The duration of the context tone was varied.

The interval between Fbs in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which listeners typically report two pitch-shift directions with equal chances). The interval between Fbs in C and T1 was fixed at 3 st and 9 st so that C would induce the maximal bias toward an upward (3 st) or a downward (9 st) pitch-shift [2]. The duration of the test

tones was 125ms each, including 2.5ms raised-cosine onset and offset ramp. There was no delay between T1 and T2. The duration of the context tone varied from trial to trial between 0ms, 5ms, 10ms, 20ms, 40ms, 80ms, 160ms and 320ms, including a 2.5ms raised-cosine onset and offset ramp. There was no delay between C and T1. Each duration condition was repeated forty times and presentation order was scrambled within each test session. Figure XX displays a schematic representation of one trial.

As in previous experiment, we sought to minimize across-trial effects by playing an inter-trial sequence of five tones between each trials. The inter-trial tones were designed as described in previous chapters ([Stimuli](#)).

Apparatus and procedure

Listeners were tested in the same conditions as described previously (see [5.1.1](#)).

Data analysis

For each listener and interval condition, the proportion of "up" responses $P(\text{up})$ was computed from the pitch-shift direction responses. As in previous experiments (Exmpt 1.1, 1.2 and 1.3), a more synthesized measure of the context effect was computed. The proportion of "biased" responses $P(\text{bias})$ was obtained by computing the proportion of time listeners responded with a bias in the direction expected from previous findings, that is with a pitch-shift encompassing the frequency region in which were situated the frequency components of the context tone [62, 2]. This way, $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones, i.e. assimilative bias whereas $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift, i.e. contrastive bias ; an absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5. The

responses times (RTs) were defined relative to the onset of T2 which correspond to the first opportunity to provide a meaningful response. Rts faster than 100ms were discarded as anticipations [233]. Because of the long-tailed distribution typical of RTs, the natural logarithms of RTs were used for all analyses [301].

6.1.2 Results

Pitch-shift direction responses

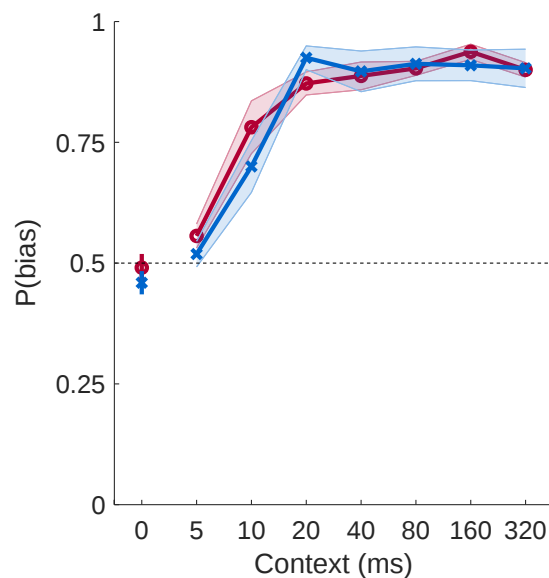


Figure 6.1.2: The thick lines represents the context bias averaged across listeners and plotted as a function of context duration conditions. In red are non-musicians and in blue, musicians. The shaded area corresponds to the standard error.

The proportion of biased response $P(\text{bias})$ averaged across individuals in each group for all context duration is displayed in Figure 6.1.2. Consistently with previous results, $P(\text{bias})$ increases with the context duration. To assess for the contribution of context duration and group on the $P(\text{bias})$, a repeated measure ANOVA with the context duration as within-subject factor and the

group as between-subject factor was conducted. It revealed that the context tone duration factor had a significant effect ($F(7, 98) = 108, p = 2e - 16$). The group factor was not significant ($F(1, 14) = 0.16, p = 0.69$) and neither was the interaction between the two factors ($F(7, 98) = 1.33, p = 0.24$).

Context (ms)	Mean context bias (SE)	<i>t</i> -test (df = 15)	<i>p</i> -value
0	0.47 (0.07)	-1.35	1.52
5	0.53 (0.07)	2.05	0.45
10	0.74 (0.15)	6.25	0.0001***
20	0.89 (0.07)	22.26	5.3e-12***
40	0.89 (0.09)	15.90	6.76e-10***
80	0.90 (0.07)	22.21	5.48e-12***
160	0.92 (0.06)	24.26	1.5e-12***
320	0.90 (0.08)	19.61	3.34e-11***

Table 6.1: The averaged context bias across all participants (both groups mixed) for each condition is compared with 0.5. The reported *p*-values are multiplied by the number of observations $N = 8$.

As in the previous experiment (Exp 1.1), we conducted a series of post-hoc *t*-tests with Bonferroni correction to evaluate which context duration was enough to have a statistically significant effect. The results of this analysis are reported in Table 6.1, the *p*-values reported are multiplied by $N = 8$).

Confidence ratings

Regarding confidence, averaged ratings across groups were plotted as a function of the interval T1-T2, as displayed in Figure 6.1.3. A repeated measure ANOVA with the context duration as within-subject factor and the group as a between-subject factor was conducted on the confidence ratings, leaving out the condition with no context. It revealed that the context duration had a significant effect ($F(6, 84) = 8.1, p = 6.41e - 07$) but not the group factor ($F(1, 14) = 0.1, p = 0.75$), nor the interaction between the two ($F(6, 84) = 1.84, p = 0.09$).

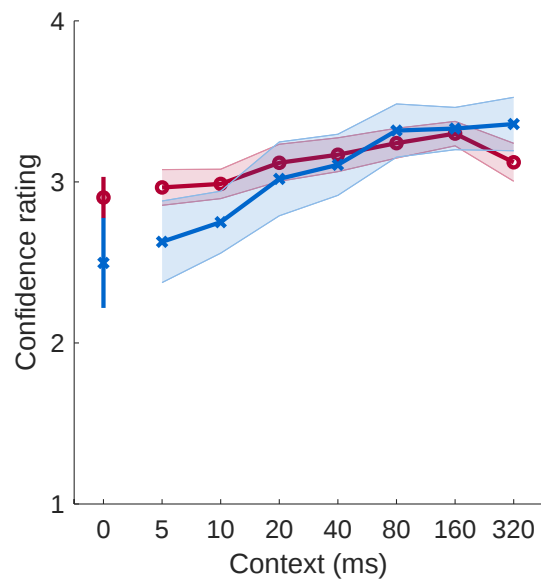


Figure 6.1.3: The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of the context duration condition (shaded areas correspond to standard error), from Z-score data (left panel) and raw data (right panel).

To observe more directly the effect of the context effect itself on confidence ratings, the correlation between the P(bias) and confidence was plotted. A linear regression was conducted on each individual data set, as displayed in Figure 6.1.4.

Each line represents the linear regression for each individuals with in blue, musicians and in red, non-musicians. We found positive correlations for almost all participants. The relation between the two variables was strong. The Pearson correlation coefficients ρ averaged across participants: $M = 0.57$, $SD = 0.36$ was statistically significant (one-sample t -test: $\rho(15) = 6.22$, $p < 0.001$).

Furthermore, a two-sample t -test was conducted on averaged coefficients in the two groups and confirmed that there was no significant difference between the groups (non-musicians: $M = 0.43$, $SD = 0.46$, musicians: $M = 0.69$, $SD = 0.15$, $t(14) = -1.5$, $p = 0.156$).

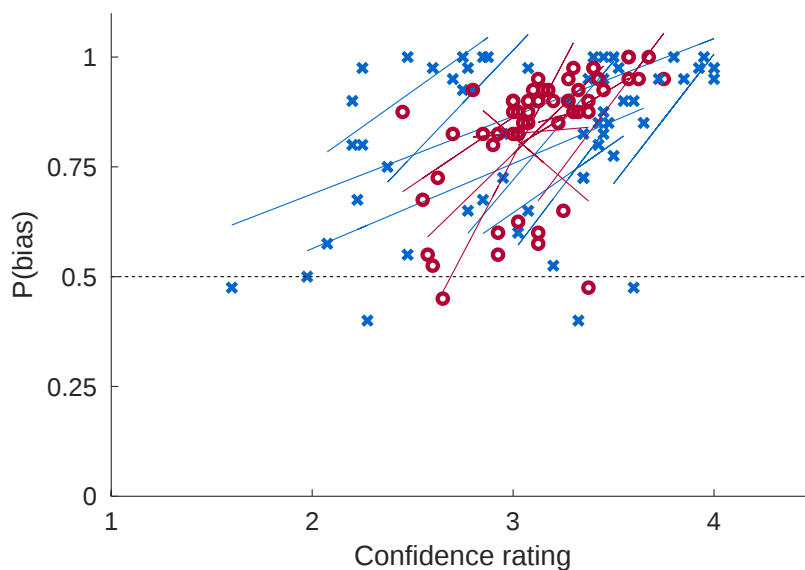


Figure 6.1.4: Individual linear regression between confidence and P(bias). In blue, individual musician data and in red, individual non-musician data.

Response time

Figure 6.1.5 displays the effect of the context duration on log-RTs, averaged across participants for the two groups, in red non-musicians and in blue, musicians.

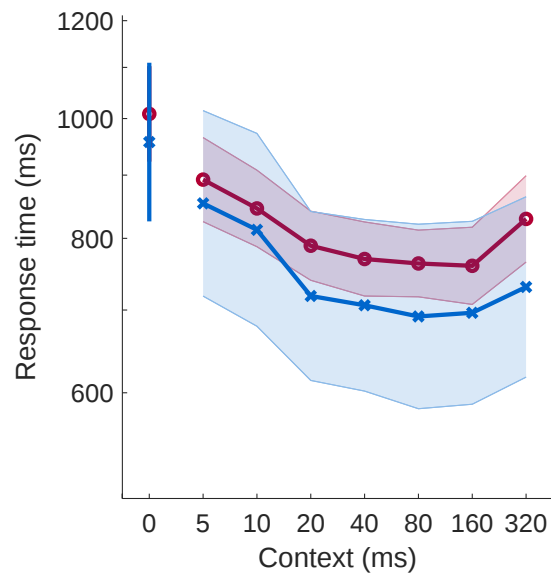


Figure 6.1.5: Averaged logarithmic response times across participants in the two groups (red: non-musicians, blue: musicians) are plotted as a function of the context duration condition (error = SE) from Z-score data (left panel) and raw data (right panel).

A repeated measure ANOVA with the context duration as a within-subject factor and the group as a between-subject factor was conducted on the log-RTs, leaving out the control condition. It revealed that the context duration had a significant effect ($F(6, 84) = 15.35, p = 8.4e - 12$) but not the group factor ($F(1, 14) = 0.2, p = 0.65$) and neither the interaction between the two ($F(1, 14) = 0.72, p = 0.63$).

The correlation between the log-RTs and the pitch-shift direction was conducted through a linear regression on each individual data set, as displayed in Figure XX. This analysis revealed that the relation between the two variables

was negative and strong. The Pearson correlation coefficients ρ averaged across participants: $M = -0.71$, $SD = 0.18$ was statistically significant ($\rho(15) = -15.57$, $p = 1e - 10$).

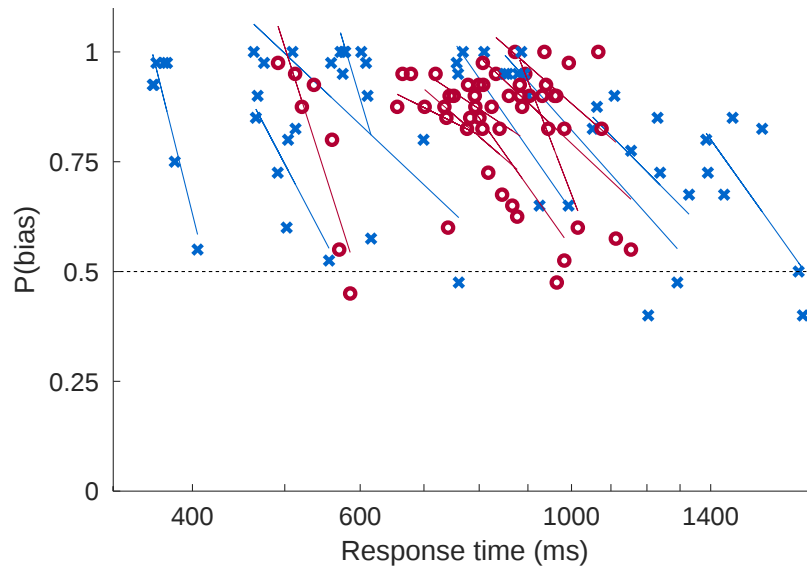


Figure 6.1.6: Top panel: Correlation between log-RTs and context bias strength. In blue, individual musician data and in red, individual non-musician data.

A two-sample t -test revealed that the two groups did not significantly differed regarding the correlation coefficient (non-musicians: $M = -0.64$, $SD = 0.19$, musicians: $M = -0.78$, $SD = 0.16$, $t(14) = 1.61$, $p = 0.129$).

6.1.3 Summary

Exp 3.1 was conducted in order to observe the effect of the context effect on the perceived ambiguity of an ambiguous auditory stimulus. In order to vary the strength of the bias induced by the context sequence, the duration of a context tone was varied. We observed an extremely fast establishment of the effect, which is consistent with previous results (Exp 1.1). However, the fact that a tone duration of only 10 ms could induce a significant bias

(vs. 20 ms in Exp 1.1) maybe due to the silence gap separating the context sequence and the test sequence (310 ms), consistently with the build up effect observed in Exp 1.3. The increase of context effect was paralleled by increasing confidence ratings and decreasing rapidity of response. In the condition with no context, non-musicians were more confident than musicians when reporting the pitch-shift of the tritone interval, consistently with previous results (however, the difference was not significant). As the context tone duration increased, the confident in both groups build up. A linear correlation was conducted on individual data and revealed that the P(bias) was directly correlated with the confidence and log-Rts.

6.2 Experiment 3.2

In Exp. 3.1, the context strength was varied by the mean of the context tone duration. In this experiment, the context strength was manipulated by varying the frequency of the context tone. We observed that the group effect disappeared when the context strength was manipulated by temporal parameters. In this experiment, we sought to test the robustness of these findings by changing the parameters used to induce the context effect. Previous findings indeed revealed that the strength of the context effect was also modulated by the interval between the context tone and the test tones, as in Figure 6.2.1 (see Appendix A and D). The perceptual bias was found to vary with the interval size. This figures indicate that the context bias displays local maxima: the maximum "upward" bias corresponds to an interval of 3st. and the maximum "downward" bias corresponds to an interval of 9st.

In this experiment, we tested the perceived ambiguity of the ambiguous interval preceded by a context sequence whose spectral parameters were manipulated to induce a stronger bias. The experimental design was:

Subj<Group2>*Context10

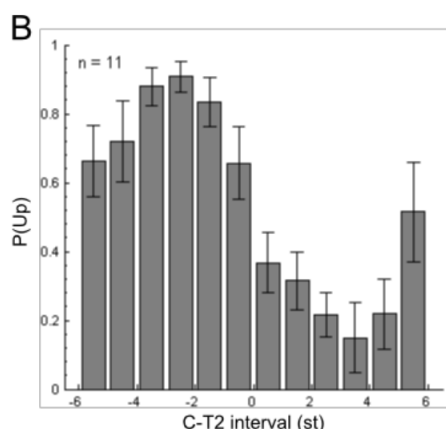


Figure 6.2.1: Results from a previous experiment (see Appendix A, section "Context Effect in Pitch Perception"). Effect of frequency of the context tone. The mean $P(T1 \text{ lower})$ is shown as a function of the interval between C and T 1 (error bars = SE).

6.2.1 Method

Participants

Sixteen self-reported normal-hearing listeners ($M = 25.6$, $SD = 5.96$), eight men and eight women, participated in the experiment. One participants had not previously taken part in experiments involving Shepard tones and three only participated to Exp 3.1. All participants passed the screening test, version 2 (see [Screening procedure](#)). Eight were musicians and eight were non musicians. The selection criteria for inclusion in the musician group was as follows: musicians had more than five years of musical practice in an academic institution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described above ([Stimuli](#)). Each trial consisted in the succession of a context sequence of one tone (C) and a test sequence composed of two test tones

(T1-T2). The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz. The interval between Fbs in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which listeners typically report two pitch-shift directions with equal chances). The interval between Fbs in C and T1 was varied between [0.75 1.5 3 4.5 5.25 6.75 7.5 9 10.5 11.25] st. The duration of the test tones was 125ms each, including 5 ms raised-cosine onset and offset ramp. There was a delay of 125 ms between T1 and T2. The duration of C was 125 ms, including a 5 ms raised-cosine onset and offset ramp. A silent gap of 500 ms was introduced between the context tone and the test tones. Each C-T1 interval condition was repeated forty times and the presentation order was shuffled within test sessions. Figure XX displays a schematic representation of one trial.

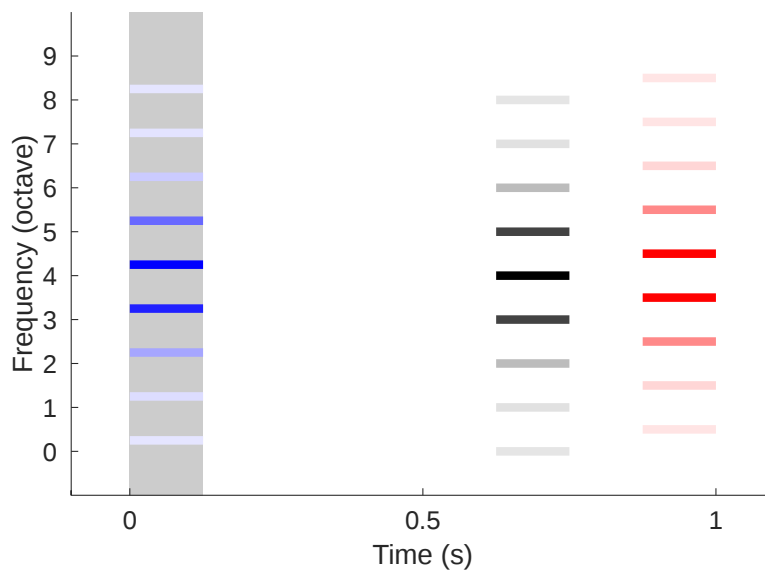


Figure 6.2.2: Schematic representation of Exp 3.2 design. The duration of the context tone was fixed and the frequency was varied between 0 st and 12 st relative to T1.

Apparatus and procedure

Listeners were tested in the same conditions as described previously (see [5.1.1](#)).

Data analysis

For each listener and interval condition, the proportion of "up" responses $P(\text{up})$ was computed from the pitch-shift direction responses. As in previous experiments (Exmpt 1.1, 1.2 and 1.3), a more synthesized measure of the context effect was computed. The proportion of "biased" responses $P(\text{bias})$ was obtained by computing the proportion of time listeners responded with a bias in the direction expected from previous findings, that is with a pitch-shift encompassing the frequency region in which were situated the frequency components of the context tone [\[62, 2\]](#). In this experiment, this corresponded to reponding "up" for [1.75 1.5 3 4.5 5.25] st and "down" for [6.75 7.5 9 10.5 11.25] st. This way, $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones, i.e. assimilative bias whereas $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift, i.e. contrastive bias ; an absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5. The responses times (RTs) were defined relative to the onset of T2 which correspond to the first opportunity to provide a meaningful response. Response times shorter than 100ms were discarded as anticipations [\[233\]](#). Because of the long-tailed distribution typical of RTs, the natural logarithms of RTs were used for all analyses [\[301\]](#).

6.2.2 Results

Pitch-shift direction

The proportion of P(bias) averaged across participants for each C-T1 interval is displayed in Figure 6.2.3. As it was observed before P(bias) increases as a function of the interval C-T1, as in Figure 6.2.1. To assess for the contribution of the interval and the group on P(bias), a repeated measure ANOVA with the frequency of the context tone as a within-subject factor and the group as a between-subject factor was conducted on the data. It revealed that the interval factor was significant ($F(4, 56) = 9.6, p = 5.29e - 6$), while the group factor was not ($F(1, 14) = 0.95, p = 0.34$). the interaction between the two factors was not significant ($F(4, 56) = 9.64, p = 0.05$).

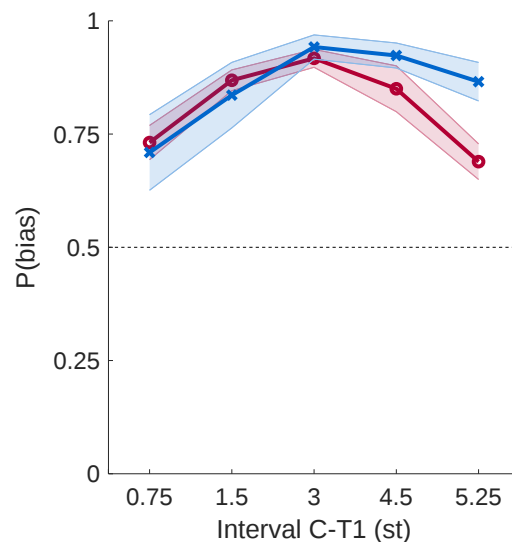


Figure 6.2.3: The thick lines represents the context bias averaged across listeners and plotted as a function of context frequency conditions. In red are non-musicians and in blue, musicians. The shaded area corresponds to the standard error.

Confidence ratings

Considering confidence ratings, Figure 6.2.4 displays the averaged confidence as a function of C-T1 interval for both groups. Interestingly, for both groups

confidence increases varies as a function of the position of C. As it is closer to the center of the frequency region, confidence ratings increase. To investigate this effect, a repeated measure ANOVA with the interval C-T1 as a within-subject factor and the group as a between-subject factor was conducted on confidence ratings. It revealed that the interval C-T1 had a significant effect ($F(4, 56) = 12.4, p = 2.52e-07$) but not the group factor ($F(1, 14) = 0.87, p = 0.36$), nor the interaction between the two ($F(4, 56) = 0.64, p = 0.63$).

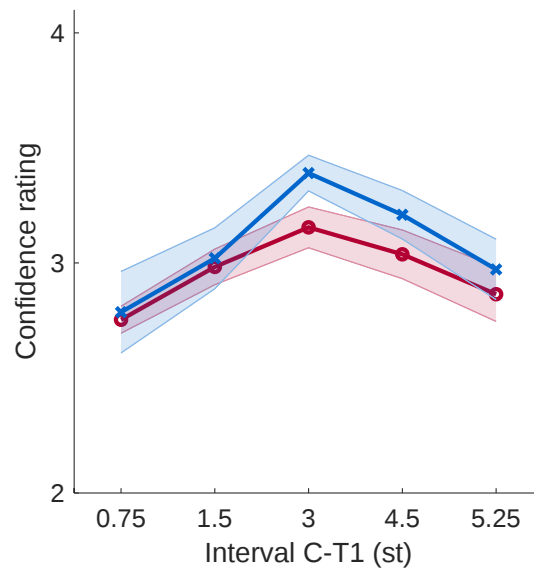


Figure 6.2.4: The thick lines represents the averaged confidence ratings across participants in the two groups (red: non-musicians, blue: musicians) plotted as a function of the context frequency condition (shaded areas correspond to standard error), from Z-score data (left panel) and raw data (right panel).

To observe more directly the effect of the context effect itself on confidence ratings, the correlation between the P(bias) and confidence was plotted. A linear regression was conducted on each individual data set, as displayed in Figure 6.2.5. Each line represents the linear regression for each individuals with in blue, musicians and in red, non-musicians. We found positive correlations for almost all participants. The relation between the two variables was strong. The Pearson correlation coefficients ρ averaged across

participants: $M = 0.66$, $SD = 0.49$ was statistically significant (one-sample t -test: $\rho(15) = 5.42$, $p < 0.001$).

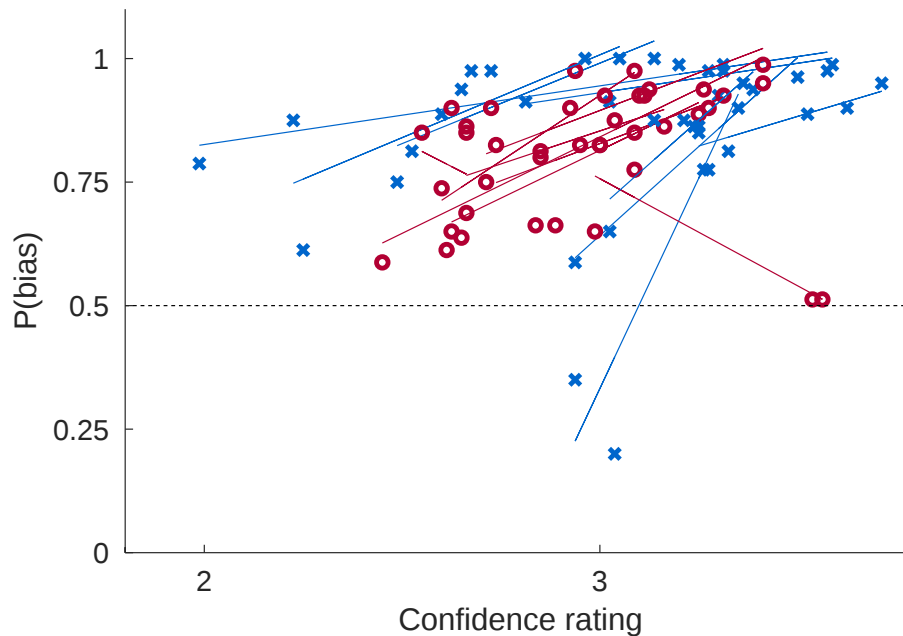


Figure 6.2.5: Individual linear regression between confidence and $P(\text{bias})$. In blue, individual musician data and in red, individual non-musician data.

Non-musicians exhibited a weaker relationship, there was no difference between groups regarding the relationship between $P(\text{bias})$ and confidence (non-musicians: $M = 0.46$, $SD = 0.64$, musicians: $M = 0.87$, $SD = 0.06$, two-sample t -test, $t(14) = -1.82$, $p = 0.09$).

Response times

Figure 6.2.6 displays the effect of the frequency of the context tone on the log-RTs, averaged across participants for the two groups. To assess for the effect of the interval C-T1 and the group factors on response times, a repeated measure ANOVA with the frequency of the context tone as a within-subject factor and the group as a between-subject factor was conducted on log-RTs.

It revealed that the interval C-T1 had a significant effect ($F(4, 56) = 9.81$, $p = 4.34e - 06$) but not the group factor ($F(1, 14) = 1.17$, $p = 0.29$) and neither the interaction between the two ($F(4, 56) = 1.29$, $p = 0.28$).

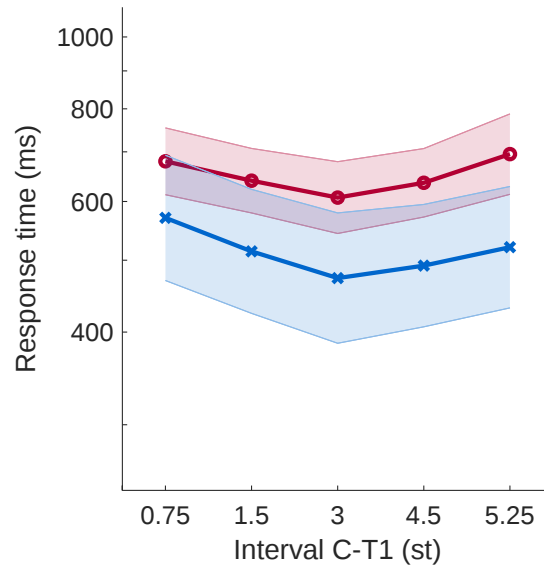


Figure 6.2.6: Averaged log-RTs across participants in the two groups (red: non-musicians, blue: musicians) are plotted as a function of the context duration condition (error = SE).

The correlation between the log-RTs and the pitch-shift direction was conducted through a linear regression on each individual data set, as displayed in Figure XX. This analysis revealed that the relation between the two variables was negative and strong. The Pearson correlation coefficients ρ averaged across participants: $M = -0.65$, $SD = 0.55$ was statistically significant ($\rho(15) = -4.69$, $p < 0.001$).

A two-sample t -test revealed that the two groups did not significantly differed regarding the correlation coefficient (non-musicians: $M = -0.44$, $SD = 0.74$, musicians: $M = -0.85$, $SD = 0.11$, $t(14) = 1.55$, $p = 0.144$).

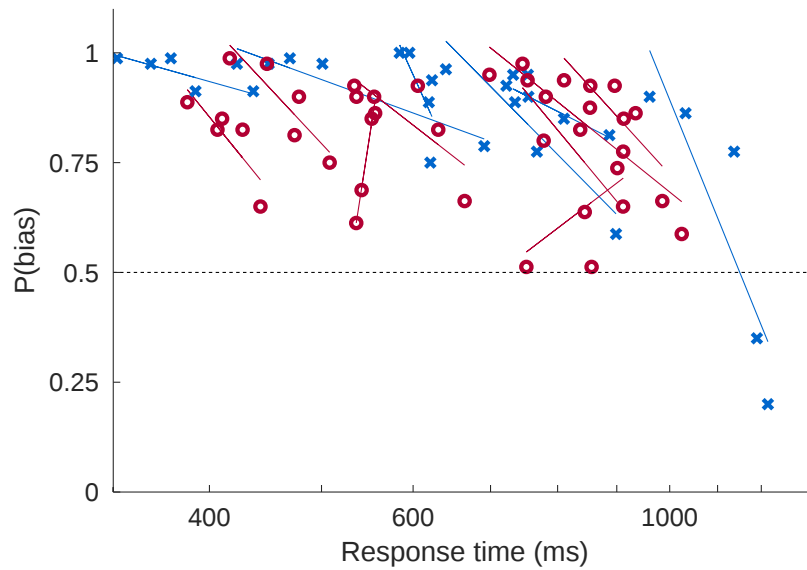


Figure 6.2.7: Individual linear regression between log-RTs and $P(\text{bias})$. In blue, individual musician data and in red, individual non-musician data.

6.2.3 Summary

In Exp 3.2, the frequency of a context tone was varied in order to manipulate the strength of the context effect. Previous findings using the same stimuli and procedures and conducted in the same laboratory demonstrated that the context effect was maximal when the frequency of the context tone was 3st. or 9st. shifted from the frequency of the test tone T1 (see Appendix A). The context bias was assessed by the direction response and the results were consistent with those previous findings. Confidence ratings and response times were measured and analyzed as a marker of the perceived ambiguity underlying the pitch-shift response. As in Exp 3.1, it was found that the strength of context, as assessed by the $P(\text{bias})$ was positively correlated with confidence ratings and negatively linked to response times. No group difference was found for both correlations.

6.3 Discussion

The two experiments presented in this chapter explore the effect of context on the ambiguity perceived by listeners associated with ambiguous pitch-shift. Previous results indicate that, when presented in isolation, an interval composed of the succession of two Shepard tones separated by a six semi-tone interval would be reported with equal probability as going "up" or "down" in pitch-shift direction [342, 90]. The ambiguous interval is sensitive to context information: when a tone is placed shortly prior the ambiguous interval, it induces a strong bias toward the shift that encompasses that tone, biasing the pitch-percept towards either an "upward" or "downward" response.

Using a combination of confidence and response time measures [303], we investigated the effect of context strength, manipulated by temporal (Exp.3.1) and spectral (Exp.3.2) parameters on the perceived ambiguity in musicians and non-musicians. Previous results have shown a strong divergence between the two groups, for ambiguous intervals with no prior context. This difference was interpreted to reflect musicians' reduced sensitivity to masking effects, thus leading them to stream out the components of Shepard tones and therefore perceive conflicting pitch-shift simultaneously. The results of two experiments revealed that the group difference disappeared as the context effect strengthened. Musicians and non-musicians were equally sensitive to the context bias and as displayed increased confidence and faster response times for stronger context bias. Individual linear regressions revealed that the bias in response and the confidence and response time measures were directly correlated. Besides, both types of context effects (temporal and spectral) led to a decrease in perceived ambiguity.

The context effect of a previous sequence of context tones is believed to reflect a process of "context-induced tracks" that biases the attributions of subsequent components heard (for more details see Appendix D), in order to optimize spectra-temporal continuity [63, 309]. In the absence of a context, the components of the Shepard tones are not attributed to determined tracks and the binding between T1 and T2 in the Shepard interval may result in

conflicting pitch-shift directions, resulting in a perceived ambiguity for musicians. However, when a context sequence is played prior to the ambiguous pair of tones, the "context-induced" tracks lead to group the components of the ambiguous interval in a way that does not lead to conflicting pitch-shifts between T1 and T2, thus reducing the perceived ambiguity.

In summary, it was shown that an ambiguous auditory stimulus [342, 90, 63] preceded by a context sequence was judged with higher confidence judgments and faster response times when it was preceded by a stronger context sequence, both triggered by temporal and spectral parameters. These findings further support the idea that contextual information is crucial when processing inconclusive sensory evidence [190]. However, the effect of context was proved to impact ambiguous perception beyond mere bias in alternative interpretation perception. The subjective experience triggered by ambiguous stimulus has started to raise interest only fairly recently and remain largely unexplored [188, 251, 371, 86, 348]. Here, it was shown that the perceived ambiguity induced by an ambiguous auditory stimulus was modulated by the bias of a prior context sequence. These findings could therefore contribute to foster the investigation of novel facets of the way context informations are exploited in the challenging task of processing highly complex natural scenes.

General discussion

When analyzing a complex scene, perceptual systems constructs a stable percept which will lead to appropriate behaviors [153, 145]. However, the information encoded by the limited range of sensors is never sufficient to specify the state of the outside world, making perception a "ill-posed problem" [97]. Resolving the ambiguity that arise from this partial set of information is therefore a crucial part of perceptual processing [179, 178]. Therefore, the myriad of processes at play when processing parsed and ambiguous sensory input into a coherent scene constitutes the core of the scientific inquiry of perception.

Whilst the visual modality has raised substantial interest among scientists, ever since the first buds of modern psychology broke, it is only fairly recently that the auditory modality started to receive a comparable attention [47].

This Ph.D. research has been dedicated to investigate auditory perception processes through the study of an ambiguous auditory object. Ambiguous stimuli were used in vision [5, 221], in audition [393, 202, 12, 255] and recently in other modalities [414, 61], in order to uncover the mechanisms at play in perceptual processes in general, for they convey maximally inconclusive sensory evidence. Shepard tones are complex harmonic tones composed of octave-related components shaped by a Gaussian envelope [342]. The circularity of Shepard tones entails interesting properties to investigate pitch perception [90]. Here, the ambiguous auditory object consists in the succession of two Shepard tones separated by a six semi-tone interval, also

referred as "tritone". This interval has been reported to result in an ambiguous percept: listeners report with an equal probabilities an "upward" or a "downward" pitch-shift [342, 90, 63, 309]. Interestingly, the pitch-shift of this six semi-tones interval being by far above the averaged threshold of normal-hearing listeners [339], the split in reported direction does not reflect uncertainty due to the difficulty of the task. Rather, it was interpreted to reflect an ambiguity in terms of temporal-spectral continuity [47, 211]: no shorter path between components may favor one or the other binding alternatives between the components of the two tones, making both pitch-shift equally likely.

The pitch-shift reported by individual listeners seems to be influenced by idiosyncratic biases [90, 91, 96, 92] (but also see [308, 93]). However, a context effect was recently reported, that rapidly overwrites idiosyncratic biases, and influence the perception of the ambiguous interval toward one or the other pitch-shift [63, 2] (see Appendix A and D). By placing a short sequence of tones before the ambiguous pair, a strong assimilative effect was observed: listeners tended to report the pitch-shift encompassing the frequency components of the context tone, with maximal efficiency for a context tone located halfway in between the tones of the ambiguous interval. This effect gradually builds up with the number of tones in the context sequence.

I first sought to characterize the temporal dynamics of this context effect, as it may be highly informative of the type of neural processes at play, as it was suggested by visual [10, 172, 312] and auditory [6, 158, 392] studies. We observed an assimilative context effect, for all tested conditions [172, 45], that was established extremely rapidly. The complementary question was to explore the remanence of the effect by introducing a silent gap of varying length between the context and the ambiguous tones. We found that the effect persisted for a surprisingly long time, 32 seconds on average (Exmpt 1.2). A third experiment was conducted to observe the interactions between the two temporal parameters and we found that the gap between the context tone and the test had an effect of enhancing the context effect (Expmpt 1.3).

Such temporal properties are remarkable, for very few context effects, either in vision [286, 172] or in audition [392, 6, 347], displayed such fast establishment properties. Similarly, the slow decay of the effect is unusual, as context effect have been observed to fade out over shorter periods of time. Slow decays have been reported in certain cases [172, 158, 352] and interpreted as markers of the fact that high-order cognitive processes are involved, the rationale being that such remanence can hardly be caused by recovery periods of sensory neurons [158]. In summary, this effect was found to be attractive with all the tested temporal parameters, suggesting that its functional relevance could be to stabilize perceptual organization [116]. Its remarkable temporal dynamics cover a wide range of time scales which suggest that the underlying mechanisms may be active in most everyday auditory scene processing, such as speech segmentation processing [183] and music listening [282].

The second part of my work was dedicated to address a simple and yet largely unanswered question: are we aware of the ambiguity? Exclusivity in perceptual ambiguity corresponds to the fact that the alternative percepts are exclusive to each other: only one is perceived at the time. It was defined as one of the hallmarks of perceptual ambiguity [221] but a growing body of evidence suggests that perceiving ambiguous stimuli can result in mixture percepts, where participants tend to report an unclear perception [157, 188, 251, 371]. A study on binocular rivalry report that mixture percepts builds up over time during the experimental session [188], the result of an anti-Hebbian plasticity [404]. After prolonged exposition of ambiguous figures, usually non-perceived in most natural scene, synaptic connection may strengthen as unwired neurons spike synchronously, as a coping strategy for processing the novel sensory input.

In the Shepard paradigm, it is commonly assumed that both percepts, the "upward" and the "downward" shift are exclusive to each other, meaning that listeners clearly hear one or the other direction, but informal comments have raised the possibility that some listeners may experience a somewhat mixture percept [90, 342, 63, 311]. The pitch-shift response was measured

for intervals with varying degrees of ambiguity [342, 63]. The six semi-tones interval was the most ambiguous case, with listeners' response split up between upward and downward responses. We also collected confidence ratings [80] and response times, as a way to measure the perceived ambiguity associated with the pitch-shift direction [371, 140]. Our results point to a strong group effect between musicians and non-musicians. Musicians tend to respond with less confidence and longer response times for the most ambiguous case and non-musicians display the opposite pattern, an "ambiguity effect" that seemingly correspond to different degrees of the perceived ambiguity (Expmt 2.1). These results were then replicated on a larger cohort in an online experiment (Expmt 2.2) and a correlation between the ambiguity effect and the number of years of musical practice was observed.

Musicianship has been associated with enhanced auditory streaming of complex scenes [411, 115, 272] and reduced sensitivity to informational masking [58, 368, 272]. In our case, we interpreted the ambiguity effect as a result of a finer attentional resolution at play in auditory scene analysis, that as been associated with expertise in the auditory [263] and visual domain [285, 139]. The Shepard tone could be streamed out into its different components by listeners with enhanced spectral resolution abilities, thus enhancing the saliency of the two percepts. Non-musicians, on the other hand, may process Shepard tones as a holistic object, without considering each component separately, which results in only one percept being salient. Interestingly, when the tones were manipulated so to increase the saliency of each component by slightly jittering their onsets, non-musicians tended to perceive the ambiguity (Expmt 2.3).

Finally, it was shown that the context effect had a similar effect on musicians and non-musicians: the bias of pitch-shift direction is paralleled by higher confidence ratings and faster response times. This was interpreted to reflect a decrease of perceived ambiguity associated with stronger context effect (Expmt 3.1 and 3.2). These findings are in line the proposed model to account for the context effect, which claims that each frequency components is assigned to a "context-induced track", based on a temporal-frequency conti-

nuity principles [47, 211]. The context constrains the grouping of frequency components in a way that favor only one or the other pitch-shift direction, leading to a decrease in the perceived ambiguity.

This Ph.D has been dedicated to the investigation of an ambiguous auditory stimulus. By exploring systematically the temporal dynamics of a context effect and the perceived ambiguity associated with the ambiguous auditory stimuli, this work has contributed to better characterize the neural processes underlying a context effect that is likely to be at play in day-to-day auditory perception. Using the same paradigm, it was also possible to explore the subjective experience associated with ambiguous stimuli. Using the combination of confidence and response measures [303, 135] in a novel setting, we demonstrated that the perception of an ambiguous stimulus was associated with different degrees of perceptual ambiguity. Despite similar perceptual responses, some listeners were sensitive to the ambiguity. From spontaneous comments, it appears that they experience a mixture of both "upward" and "downward" pitch-shift, as it has been reported in other studies on ambiguous stimuli [188, 86, 251] although with various frequency [348, 295]. Here, we found that various degree of perceived ambiguity was correlated with musical expertise. This findings therefore further support the hypothesis that musicians have enhanced abilities in auditory scene analysis related tasks [58, 368, 272, 411, 115], a advantage that could be attributed to the specificity of musical training in parsing very complex acoustic scenes such as an orchestra ensemble [296, 263].

In my view, the most interesting aspect of this research pertains to the distinction between vague and ambiguous perception that can be sketched. In a special issue paper on perceptual ambiguity, Schwartz et al. suggest that exclusivity is a specificity of ambiguous stimuli carrying conflicting binding cues, such as the duck-rabbit or the Rubin face/vase: for each perceptual alternative to be perceived, an overall reorganization of the stimulus must be operated, which implies that both percepts can not be perceived simultaneously [334]. The perception of vague stimuli, on the other hand, is characterized by the fact that response categories are fuzzy, non-exclusive, and observers

are aware of their uncertainty when selecting a response [299, 105, 140]. Interestingly, the binding organization of perceptual objects can be modulated by different aspects [213] such as the synchronicity of its components [276] or prior knowledge [316, 134]. Vagueness and ambiguity experiences may therefore be triggered by factors extraneous to the signal itself, such as prior information or a priori knowledge.

In that framework, confidence ratings and response times could arguably constitute well-fitted tools to disentangle perception of vague and ambiguous stimuli, as the difference lies in introspective aspects of the perceptual experience that can be addressed with those measures. In an experiment investigating the processes of perceptual categorization, longer response times were associated with categorical indecision [140, 111]. In contrast, diverse degrees of ambiguity in a ambiguous motion were found to cause no effect on response times [371]. In that framework, our findings suggest that musicians experienced the interval as vague, as both percepts were available simultaneously, resulting in lower confidence ratings and slower response times. In contrast, non-musicians displayed a behavior indicative of perceptual ambiguity: only one percept at a time was experienced, resulting in high confidence ratings and fast response times. It was hypothesized that the specificity of musicians consisted in an enhanced spectral resolution on the sub-parts of the scene, a consequence of expertise [285, 138, 139] in auditory scene analysis tasks [296]. It may also be argued that the difference between musicians and non-musicians lies into their a priori knowledge of sound organizations, that may result in different perceptual experiences of the ambiguity [134, 213]. Musicians, being trained to interact with extremely complex auditory scenes -for instance when they play in an orchestra-, develop a stronger knowledge of the fact that even synchronous sounds may originate from distinct sources [296]. This could account for the fact that they maintain a high attention to the distinct components of the Shepard tones, considering that they potentially originate from distinct sources. Strikingly, when non-musicians have access to cues that enhance the streaming of Shepard tones into its different components, they perceive the ambiguity as their attention is drawn

to sub-parts of the scene. In conclusion, this research contributed to develop a novel method to explore introspective aspects of the perception of inconclusive stimuli. By shedding light on a relevant distinction between vagueness and ambiguity, this work could therefore foster new perspectives on the processes as play when dealing with complex auditory scenes.

Bibliography

- [1] M. Aafjes, J. E. Hueting, and P. Visser. Individual and interindividual differences in binocular retinal rivalry in man. *Psychophysiology*, 3(1):18–22, 1966.
- [2] V. Adam, M. Sahani, C. Pelofi, C. Chambers, and D. Pressnitzer. *Pre-perceptual grouping of auditory scenes explains contextual biases in the perception of ambiguous tonal shifts*. July 2015.
- [3] E. D. Adrian. The basis of sensation. 1928.
- [4] N. C. Aggelopoulos. Perceptual inference. *Neuroscience & Biobehavioral Reviews*, 2015.
- [5] D. Alais. Binocular rivalry: competition and inhibition in visual perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1):87–103, 2012.
- [6] D. Alais, E. Orchard-Mills, and E. Van der Burg. Auditory frequency perception adapts rapidly to the immediate past. *Attention, Perception, & Psychophysics*, 77(3):896–906, 2015.
- [7] S. Alxatib and F. J. Pelletier. The psychology of vagueness: Borderline cases and contradictions. *Mind & Language*, 26(3):287–326, 2011.
- [8] O. Amir, N. Amir, and L. Kishon-Rabin. The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America*, 113(2):1102–1108, 2003.

- [9] T. J. Andrews and C. Blakemore. Integration of motion information during binocular rivalry. *Vision research*, 42(3):301–309, 2002.
- [10] S. Anstis and V. S. Ramachandran. Visual inertia in apparent motion. *Vision research*, 27(5):755–764, 1987.
- [11] S. Anstis, F. A. Verstraten, and G. Mather. The motion aftereffect. *Trends in cognitive sciences*, 2(3):111–117, 1998.
- [12] S. M. Anstis and S. Saida. Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, 11(3):257, 1985.
- [13] P. Ashourian and Y. Loewenstein. Bayesian inference underlies the contraction bias in delayed comparison tasks. *PloS one*, 6(5):e19551, 2011.
- [14] V. M. Bajo, F. R. Nodal, D. R. Moore, and A. J. King. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature neuroscience*, 13(2):253–260, 2010.
- [15] A. Basirat, M. Sato, J.-L. Schwartz, P. Kahane, and J.-P. Lachaux. Parieto-frontal gamma band activity during the perceptual emergence of speech forms. *NeuroImage*, 42(1):404–413, 2008.
- [16] D. Başkent and E. Gaudrain. Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, 139(3):EL51–EL56, 2016.
- [17] M. W. Beauvois and R. Meddis. Time decay of auditory stream biasing. *Perception & Psychophysics*, 59(1):81–86, Jan. 1997.
- [18] S. L. Bengtsson, Z. Nagy, S. Skare, L. Forsman, H. Forssberg, and F. Ullén. Extensive piano practicing has regionally specific effects on white matter development. *Nature neuroscience*, 8(9):1148–1150, 2005.

- [19] P. Bermudez and R. J. Zatorre. Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*, 1060(1):395–399, 2005.
- [20] C. Bey and S. McAdams. Schema-based processing in auditory scene analysis. *Perception & psychophysics*, 64(5):844–854, 2002.
- [21] J. J. Bharucha. Music cognition and perceptual facilitation: A connectionist framework. *Music Perception: An Interdisciplinary Journal*, 5(1):1–30, 1987.
- [22] J. J. Bharucha and K. Stoeckig. Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, 12(4):403, 1986.
- [23] J. J. Bharucha and K. Stoeckig. Priming of chords: spreading activation or overlapping frequency spectra? *Perception & Psychophysics*, 41(6):519–524, 1987.
- [24] G. M. Bidelman and A. Krishnan. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research*, 1355:112–125, 2010.
- [25] G. M. Bidelman, J. M. Schug, S. G. Jennings, and S. P. Bhagat. Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. *The Journal of the Acoustical Society of America*, 136(1):EL33–EL39, 2014.
- [26] E. Bigand. Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music*, 18(1):45–59, 1990.
- [27] E. Bigand. Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3):808, 1997.
- [28] E. Bigand, F. Madurell, B. Tillmann, and M. Pineau. Effect of global structure and temporal organization on chord processing. *Journal*

- of *Experimental Psychology: Human Perception and Performance*, 25(1):184, 1999.
- [29] E. Bigand and R. Parncutt. Perceiving musical tension in long chord sequences. *Psychological Research*, 62(4):237–254, 1999.
- [30] E. Bigand, R. Parncutt, and F. Lerdahl. Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, 58(1):125–141, 1996.
- [31] E. Bigand and M. Pineau. Global context effects on musical expectancy. *Perception & Psychophysics*, 59(7):1098–1107, 1997.
- [32] E. Bigand, B. Poulin, B. Tillmann, F. Madurell, and D. A. D'Adamo. Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human perception and performance*, 29(1):159, 2003.
- [33] E. Bigand and B. Poulin-Charronnat. Are we experienced listeners? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100(1):100–130, 2006.
- [34] E. Bigand, B. Tillmann, B. Poulin, D. A. D'Adamo, and F. Madurell. The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, 81(1):B11–B20, 2001.
- [35] E. Bigand, B. Tillmann, B. Poulin-Charronnat, and D. Manderlier. Repetition priming: Is music special? *The Quarterly Journal of Experimental Psychology Section A*, 58(8):1347–1375, 2005.
- [36] J. Blacking. *How musical is man?* University of Washington Press, 1974.
- [37] R. Blake. A neural theory of binocular rivalry. *Psychological review*, 96(1):145, 1989.

- [38] T. V. Bliss, G. L. Collingridge, and others. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361(6407):31–39, 1993.
- [39] D. Boebinger, S. Evans, S. Rosen, C. F. Lima, T. Manly, and S. K. Scott. Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, 137(1):378–387, 2015.
- [40] A. Bonneville-Roussy and T. Bouffard. When quantity is not enough: Disentangling the roles of practice time, self-regulation and deliberate practice in musical achievement. *Psychology of Music*, page 0305735614534910, 2014.
- [41] H. A. Bouchey and S. Harter. Reflected Appraisals, Academic Self-Perceptions, and Math/Science Performance During Early Adolescence. *Journal of Educational Psychology*, 97(4):673, 2005.
- [42] D. H. Brainard and A. C. Hurlbert. Colour vision: understanding#TheDress. *Current Biology*, 25(13):R551–R554, 2015.
- [43] J. W. Brascamp, P. C. Klink, and W. J. M. Levelt. The ‘laws’ of binocular rivalry: 50 years of Levelt’s propositions. *Vision Research*, 109, Part A:20–37, 2015.
- [44] J. W. Brascamp, T. H. Knapen, R. Kanai, A. J. Noest, R. Van Ee, and A. V. Van Den Berg. Multi-timescale perceptual history resolves visual ambiguity. *PloS one*, 3(1):e1497, 2008.
- [45] J. W. Brascamp, T. H. Knapen, R. Kanai, R. Van Ee, and A. V. Van Den Berg. Flash suppression and flash facilitation in binocular rivalry. *Journal of vision*, 7(12):12, 2007.
- [46] J. W. Brascamp, J. Pearson, R. Blake, and A. V. Van Den Berg. Intermittent ambiguous stimuli: implicit memory causes periodic perceptual alternations. *Journal of Vision*, 9(3):3–3, 2009.

- [47] A. S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, 1994.
- [48] J. Britz, T. Landis, and C. M. Michel. Right parietal brain activity precedes perceptual alternation of bistable stimuli. *Cerebral Cortex*, 19(1):55–65, 2009.
- [49] M. Brosch and C. E. Schreiner. Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, 77(2):923–943, 1997.
- [50] G. J. Brouwer and R. van Ee. Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Vision Research*, 46(20):3393–3402, Oct. 2006.
- [51] J. S. Bruner and A. L. Minturn. Perceptual identification and perceptual organization. *The Journal of General Psychology*, 53(1):21–28, 1955.
- [52] H. Bulf, S. P. Johnson, and E. Valenza. Visual statistical learning in the newborn infant. *Cognition*, 121(1):127–132, 2011.
- [53] E. M. Burns. Circularity in relative pitch judgments for inharmonic complex tones: The Shepard demonstration revisited, again. *Perception & Psychophysics*, 30(5):467–472, 1981.
- [54] E. M. Burns and W. D. Ward. Categorical perception of musical intervals. *The Journal of the Acoustical Society of America*, 55(2):456–456, 1974.
- [55] E. M. Burns and W. D. Ward. Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, 63(2):456–468, Feb. 1978.
- [56] M. B. Calford and M. N. Semple. Monaural inhibition in cat auditory cortex. *Journal of Neurophysiology*, 73(5):1876–1891, 1995.

- [57] G. V. Caprara, R. Fida, M. Vecchione, G. Del Bove, G. M. Vecchio, C. Barbaranelli, and A. Bandura. Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of educational psychology*, 100(3):525, 2008.
- [58] D. Carey, S. Rosen, S. Krishnan, M. T. Pearce, A. Shepherd, J. Aydelott, and F. Dick. Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, 137:81–105, 2015.
- [59] P. A. Cariani and B. Delgutte. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76(3):1698–1716, 1996.
- [60] R. P. Carlyon and R. Cusack. Effects of attention on auditory perceptual organisation. *Neurobiology of Attention*. Academic Press, San Diego, pages 317–323, 2005.
- [61] O. Carter, T. Konkle, Q. Wang, V. Hayward, and C. Moore. Tactile rivalry demonstrated with an ambiguous apparent-motion quartet. *Current Biology*, 18(14):1050–1054, 2008.
- [62] C. Chambers, S. Akram, V. Adam, C. Pelofi, M. Sahani, S. Shamma, and D. Pressnitzer. Prior context in audition informs binding and shapes simple features (under review). *Nature Communications*.
- [63] C. Chambers and D. Pressnitzer. Perceptual hysteresis in the judgment of auditory pitch shift. *Attention, Perception, & Psychophysics*, 76(5):1271–1279, May 2014.
- [64] C. D. Chambers, J. B. Mattingley, and S. A. Moss. The octave illusion revisited: Suppression or fusion between ears? *Journal of Experimental Psychology: Human Perception and Performance*, 28(6):1288, 2002.
- [65] T. Chin and N. S. Rickard. The music USE (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception*, 29(4):429–446, 2012.

- [66] S. C. Chong, D. Tadin, and R. Blake. Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of vision*, 5(11):6, 2005.
- [67] K. L. Christison-Lagay, A. M. Gifford, and Y. E. Cohen. Neural correlates of auditory scene analysis and perception. *International journal of psychophysiology*, 95(2):238–245, 2015.
- [68] T. Collins, B. Tillmann, F. S. Barrett, C. Delbé, and P. Janata. A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior. *Psychological review*, 121(1):33, 2014.
- [69] C. D. Condon and N. M. Weinberger. Habituation produces frequency-specific plasticity of receptive fields in the auditory cortex. *Behavioral neuroscience*, 105(3):416, 1991.
- [70] V. Conrad, A. Bartels, M. Kleiner, and U. Noppeney. Audiovisual interactions in binocular rivalry. *Journal of Vision*, 10(10):27, 2010.
- [71] M. Corbetta and G. L. Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3):201–215, 2002.
- [72] N. Cowan. On short and long auditory stores. *Psychological bulletin*, 96(2):341, 1984.
- [73] I. Cross. Music as a biocultural phenomenon. *Annals of the New York Academy of Sciences*, 999(1):106–111, 2003.
- [74] R. Cusack. The intraparietal sulcus and perceptual organization. *Journal of cognitive neuroscience*, 17(4):641–651, 2005.
- [75] J. C. Dahmen, P. Keating, F. R. Nodal, A. L. Schulz, and A. J. King. Adaptation to stimulus statistics in the perception and neural representation of auditory space. *Neuron*, 66(6):937–948, 2010.

- [76] J. C. Dahmen and A. J. King. Learning to hear: plasticity of auditory cortical processing. *Current opinion in neurobiology*, 17(4):456–464, 2007.
- [77] G. D. Davidson and M. A. Pitts. Auditory event-related potentials associated with perceptual reversals of bistable pitch motion. *Frontiers in Human Neuroscience*, 8, Aug. 2014.
- [78] A. De Cheveigne. Pitch perception models. In *Pitch*, pages 169–233. Springer, 2005.
- [79] V. De Gardelle, L. Charles, and S. Kouider. Perceptual awareness and categorical representation of faces: Evidence from masked priming. *Consciousness and cognition*, 20(4):1272–1281, 2011.
- [80] V. De Gardelle and P. Mamassian. Weighting mean and variability during confidence judgments. *PloS one*, 10(3):e0120870, 2015.
- [81] S. Deike, P. Heil, M. Böckmann-Barthel, and A. Brechmann. The build-up of auditory stream segregation: a different perspective. *Frontiers in psychology*, 3, 2012.
- [82] L. Demany. Auditory stream segregation in infancy. *Infant Behavior and Development*, 5(2–4):261–276, 1982.
- [83] L. Demany, D. Pressnitzer, and C. Semal. Tuning properties of the auditory frequency-shift detectors. *The Journal of the Acoustical Society of America*, 126(3):1342–1348, 2009.
- [84] L. Demany and C. Ramos. On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *The Journal of the Acoustical Society of America*, 117(2):833–841, 2005.
- [85] L. Demany, C. Semal, and D. Pressnitzer. Implicit versus explicit frequency comparisons: Two mechanisms of auditory change detection. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2):597, 2011.

- [86] S. Denham, T. M. Böhm, A. Bendixen, O. Szalárdy, Z. Kocsis, R. Mill, and I. Winkler. Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Frontiers in neuroscience*, 8, 2014.
- [87] S. L. Denham, K. Gyimesi, G. Stefanics, and I. Winkler. Perceptual bistability in auditory streaming: how much do stimulus features matter? *Learning & Perception*, 5(Supplement 2):73–100, 2013.
- [88] D. Deutsch. Music recognition. *Psychological Review*, 76(3):300, 1969.
- [89] D. Deutsch. An auditory illusion. *The Journal of the Acoustical Society of America*, 55(S1):S18–S19, 1974.
- [90] D. Deutsch. A musical paradox. *Music Perception*, pages 275–280, 1986.
- [91] D. Deutsch. The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, 41(6):563–575, 1987.
- [92] D. Deutsch. The tritone paradox: An influence of language on music perception. *Music Perception: An Interdisciplinary Journal*, 8(4):335–347, 1991.
- [93] D. Deutsch. The tritone paradox and the pitch range of the speaking voice: Reply to Repp. *Music Perception*, pages 257–263, 1994.
- [94] D. Deutsch, T. Henthorn, and M. Dolson. Speech patterns heard early in life influence later perception of the tritone paradox. 2004.
- [95] D. Deutsch, W. L. Kuyper, and Y. Fisher. The tritone paradox: Its presence and form of distribution in a general population. *Music Perception*, pages 79–92, 1987.
- [96] D. Deutsch, T. North, and L. Ray. The tritone paradox: Correlate with the listener’s vocal range for speech. *Music Perception*, pages 371–384, 1990.

- [97] V. Di Lollo. The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, 16(6):317–321, 2012.
- [98] T. Ditzinger, B. Tuller, and J. S. Kelso. Temporal patterning in an auditory illusion: the verbal transformation effect. *Biological cybernetics*, 77(1):23–30, 1997.
- [99] D. W. Dong and J. J. Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358, 1995.
- [100] W. J. Dowling. Chapter Fifteen Perception of Music. 2001.
- [101] W. R. Drennan, S. Gatehouse, and C. Lever. Perceptual segregation of competing speech sounds: the role of spatial location. *The Journal of the Acoustical Society of America*, 114(4):2178–2189, 2003.
- [102] N. I. Durlach, C. R. Mason, G. Kidd Jr, T. L. Arbogast, H. S. Colburn, and B. G. Shinn-Cunningham. Note on informational masking (L). *The Journal of the Acoustical Society of America*, 113(6):2984–2987, 2003.
- [103] E. F. Dutour. *Discussion d’une question d’optique*. De l’Imprimerie Royale, 1760.
- [104] P. Egré. Soritical series and Fisher series. In *Reduction between the mind and the brain*, pages 91–115. eds H. Leitgeb & A. Hieke, heusenstamm, germany: ontos-verlag edition, 2009.
- [105] P. Egré, V. De Gardelle, and D. Ripley. Vagueness and order effects in color categorization. *Journal of Logic, Language and Information*, 22(4):391–420, 2013.
- [106] J. Ehlers, D. Strüber, and C. Basar-Eroglu. Multistable perception in children: Prefrontal delta oscillations in the developing brain. *International Journal of Psychophysiology*, 2015.
- [107] P. D. Eimas, E. R. Siqueland, P. Jusczyk, and J. Vigorito. Speech perception in infants. *Science*, 171(3968):303–306, 1971.

BIBLIOGRAPHY

- [108] T. Elbert, C. Pantev, C. Wienbruch, B. Rockstroh, and E. Taub. Increased cortical representation of the fingers of the left hand in string players. *Science*, 270(5234):305–307, 1995.
- [109] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3):363, 1993.
- [110] M. O. Ernst and H. H. Bühlhoff. Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169, 2004.
- [111] B. Espinoza-Varas and C. S. Watson. Effects of decision criterion on response latencies of binary decisions. *Perception & Psychophysics*, 55(2):190–203, 1994.
- [112] G. Fairbanks and P. Grubb. A psychophysical investigation of vowel formants. *Journal of Speech & Hearing Research*, 1961.
- [113] F. Fang and S. He. Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature neuroscience*, 8(10):1380–1385, 2005.
- [114] R. R. Fay. Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*). *Journal of the Association for Research in Otolaryngology*, 1(2):120–128, 2000.
- [115] P. A. Fine and B. C. Moore. Frequency analysis and musical ability. *Music Perception: An Interdisciplinary Journal*, 11(1):39–53, 1993.
- [116] J. Fischer and D. Whitney. Serial dependence in visual perception. *Nature Neuroscience*, 17(5):738–743, 2014.
- [117] Y. I. Fishman, D. H. Reser, J. C. Arezzo, and M. Steinschneider. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing research*, 151(1):167–187, 2001.

- [118] S. M. Fleming and H. C. Lau. How to measure metacognition. *Frontiers in human neuroscience*, 8(July):1–9, 2014.
- [119] S. M. Fleming, R. S. Weil, Z. Nagy, R. J. Dolan, and G. Rees. Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998):1541–1543, 2010.
- [120] C. A. Fowler. Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68(2):161–177, 2006.
- [121] P. Fraisse. Visual perceptive simultaneity and masking of letters successively presented. *Perception & Psychophysics*, 1(9):285–287, 1966.
- [122] E. D. Freeman and J. Driver. Subjective appearance of ambiguous structure-from-motion can be driven by objective switches of a separate less ambiguous context. *Vision research*, 46(23):4007–4023, 2006.
- [123] T. Fujioka, L. J. Trainor, B. Ross, R. Kakigi, and C. Pantev. Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of cognitive neuroscience*, 16(6):1010–1021, 2004.
- [124] T. Fujioka, L. J. Trainor, B. Ross, R. Kakigi, and C. Pantev. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of cognitive neuroscience*, 17(10):1578–1592, 2005.
- [125] J. M. Fuster. Frontal lobe and cognitive development. *Journal of Neurocytology*, 31(3-5):373–385, 2002.
- [126] R. B. Gardner and J. P. Wilson. Evidence for direction-specific channels in the processing of frequency modulation. *The Journal of the Acoustical Society of America*, 66(3):704–709, 1979.
- [127] E. Genç, J. Bergmann, W. Singer, and A. Kohler. Interhemispheric connections shape subjective experience of bistable motion. *Current Biology*, 21(17):1494–1499, 2011.

- [128] S. Gepshtein and M. Kubovy. Stability and change in perception: spatial organization in temporal context. *Experimental brain research*, 160(4):487–495, 2005.
- [129] S. Getzmann and R. Näätänen. The mismatch negativity as a measure of auditory stream segregation in a simulated “cocktail-party” scenario: effect of age. *Neurobiology of Aging*, 2015.
- [130] Z. Ghahramani and M. I. Jordan. Factorial hidden Markov models. *Machine learning*, 29(2-3):245–273, 1997.
- [131] J. Giangrand, B. Tuller, and J. S. Kelso. Perceptual dynamics of circular pitch. *Music Perception*, 20(3):241–262, 2003.
- [132] J. Giangrande. The tritone paradox: Effects of pitch class and position of the spectral envelope. *Music Perception*, pages 253–264, 1998.
- [133] J. N. Giedd, J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A. C. Evans, and J. L. Rapoport. Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, 2(10):861–863, Oct. 1999.
- [134] J. J. Girgus, I. Rock, and R. Egatz. The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception & Psychophysics*, 22(6):550–556, 1977.
- [135] A. Glöckner and A. Bröder. Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making*, 6(1):23, 2011.
- [136] P. Goolkasian. Ambiguous figures: role of context and critical features. *Journal of General Psychology*, 1987.
- [137] E. W. Graf and W. J. Adams. Surface organization influences bistable vision. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2):502, 2008.

- [138] C. S. Green and D. Bavelier. Action-video-game experience alters the spatial resolution of vision. *Psychological science*, 18(1):88–94, 2007.
- [139] C. S. Green and D. Bavelier. Learning, attentional control, and action video games. *Current Biology*, 22(6):R197–R206, 2012.
- [140] J. Grinband, J. Hirsch, and V. P. Ferrera. A neural representation of categorization uncertainty in the human brain. *Neuron*, 49(5):757–763, 2006.
- [141] B. Grothe, M. Pecka, and D. McAlpine. Mechanisms of Sound Localization in Mammals. *Physiological Reviews*, 90(3):983–1012, July 2010.
- [142] M. Grube, D. Y. von Cramon, and R. Rübsamen. Inharmonicity detection. *Experimental brain research*, 153(4):637–642, 2003.
- [143] A. Gutschalk and A. R. Dykstra. Functional imaging of auditory scene analysis. *Hearing research*, 307:98–110, 2014.
- [144] A. Gutschalk, C. Micheyl, J. R. Melcher, A. Rupp, M. Scherg, and A. J. Oxenham. Neuromagnetic correlates of streaming in human auditory cortex. *The Journal of Neuroscience*, 25(22):5382–5388, 2005.
- [145] P. Haazebroek, S. Van Dantzig, and B. Hommel. How task goals mediate the interplay between perception and action. 2013.
- [146] M. Habib and M. Besson. What do music training and musical experience teach us about brain plasticity? 2009.
- [147] S. Hallam. 21st century conceptions of musical ability. *Psychology of Music*, 38(3):308–330, 2010.
- [148] S. Harnad. Categorical perception. *Encyclopedia of cognitive science*, 67(4), 2003.

BIBLIOGRAPHY

- [149] J.-D. Haynes, R. Deichmann, and G. Rees. Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature*, 438(7067):496–499, 2005.
- [150] D. O. Hebb. *The organization of behavior*. New York: Wiley, 1949.
- [151] H. v. Helmholtz and J. P. C. Southall. *Treatise on Physiological Optics*. Courier Corporation, 2005.
- [152] C. M. Hempel, K. H. Hartman, X.-J. Wang, G. G. Turrigiano, and S. B. Nelson. Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *Journal of neurophysiology*, 83(5):3031–3041, 2000.
- [153] A. Herwig. Linking perception and action by structure or process? Toward an integrative perspective. *Neuroscience & Biobehavioral Reviews*, 52:105–116, 2015.
- [154] K. T. Hill, C. W. Bishop, and L. M. Miller. Auditory grouping mechanisms reflect a sound’s relative position in a sequence. *Frontiers in human neuroscience*, 6, 2012.
- [155] K. T. Hill, C. W. Bishop, D. Yadav, and L. M. Miller. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC neuroscience*, 12(1):85, 2011.
- [156] H. L. Hollingworth. The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, pages 461–469, 1910.
- [157] M. Hollins. The effect of contrast on the completeness of binocular rivalry suppression. *Perception & psychophysics*, 27(6):550–556, 1980.
- [158] L. L. Holt. Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, 16(4):305–312, 2005.
- [159] L. L. Holt and A. J. Lotto. Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167(1):156–169, 2002.

- [160] L. L. Holt and A. J. Lotto. Speech perception within an auditory cognitive science framework. *Current directions in psychological science*, 17(1):42–46, 2008.
- [161] L. L. Holt, A. J. Lotto, and K. R. Kluender. Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America*, 108(2):710–722, 2000.
- [162] S. H. Hulse, S. A. MacDougall-Shackleton, and A. B. Wisniewski. Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 111(1):3, 1997.
- [163] J.-M. Hupe, L.-M. Joffo, and D. Pressnitzer. Bistability for audiovisual stimuli: Perceptual decision is modality specific. *Journal of Vision*, 8(7):1–1, May 2008.
- [164] D. B. Huron. *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.
- [165] M. Intaité, V. Noreika, A. Soliunas, and C. M. Falter. Interaction of bottom-up and top-down processes in the perception of ambiguous figures. *Vision Research*, 89:24–31, Aug. 2013.
- [166] M. Jazayeri and M. N. Shadlen. Temporal context calibrates interval timing. *Nature neuroscience*, 13(8):1020–1026, 2010.
- [167] S. Jentschke, S. Koelsch, and A. D. Friederici. Investigating the relationship of music and language in children. *Annals of the New York Academy of Sciences*, 1060(1):231–242, 2005.
- [168] J. P. Kalisvaart, I. Klaver, and J. Goossens. Motion discrimination under uncertainty and ambiguity. *Journal of vision*, 11(1):20, 2011.
- [169] R. Kanai, B. Bahrami, and G. Rees. Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current biology*, 20(18):1626–1630, 2010.

- [170] R. Kanai, D. Carmel, B. Bahrami, and G. Rees. Structural and functional fractionation of right superior parietal cortex in bistable perception. *Current biology*, 21(3):R106–R107, 2011.
- [171] R. Kanai and G. Rees. The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, 12(4):231–242, 2011.
- [172] R. Kanai and F. A. Verstraten. Perceptual manifestations of fast neural plasticity: Motion priming, rapid motion aftereffect and perceptual sensitization. *Vision research*, 45(25):3109–3116, 2005.
- [173] R. Kanai and F. A. Verstraten. Attentional modulation of perceptual stabilization. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1591):1217–1222, 2006.
- [174] M. Kashino and H. M. Kondo. Functional brain networks underlying perceptual switching: auditory streaming and verbal transformations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):977–987, May 2012.
- [175] M. Kashino and S. Nishida. Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *The Journal of the Acoustical Society of America*, 103(6):3597–3604, 1998.
- [176] T. Kayahara. Aftereffect of adaptation to uni-directional frequency change: Evidence for selective processing mechanism. *Acoustical Science and Technology*, 22(1):49–51, 2001.
- [177] J. P. Keenan, V. Thangaraj, A. R. Halpern, and G. Schlaug. Absolute pitch and planum temporale. *Neuroimage*, 14(6):1402–1408, 2001.
- [178] D. Kersten, P. Mamassian, and A. Yuille. Object Perception as Bayesian Inference. *Annual Review of Psychology*, 55(1):271–304, 2004.

- [179] D. Kersten and A. Yuille. Bayesian models of object perception. *Current opinion in neurobiology*, 13(2):150–158, 2003.
- [180] G. Kidd Jr, C. R. Mason, V. M. Richards, F. J. Gallun, and N. I. Durlach. Informational masking. In *Auditory perception of sound sources*, pages 143–189. Springer, 2008.
- [181] J. A. Kiernan. *Introduction to human neuroscience*. Lippincott Williams & Wilkins, 1987.
- [182] L. Kishon-Rabin, O. Amir, Y. Vexler, and Y. Zaltz. Pitch discrimination: Are professional musicians better than non-musicians? *Journal of basic and clinical physiology and pharmacology*, 12(2):125–144, 2001.
- [183] D. H. Klatt. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221, 1976.
- [184] D. A. Kleffner and V. S. Ramachandran. On the perception of shape from shading. *Perception & Psychophysics*, 52(1):18–36, 1992.
- [185] A. Kleinschmidt, C. Büchel, C. Hutton, K. J. Friston, and R. S. Frackowiak. The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron*, 34(4):659–666, 2002.
- [186] A. Kleinschmidt, C. Büchel, S. Zeki, and R. S. J. Frackowiak. Human brain activity during spontaneously reversing perception of ambiguous figures. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1413):2427–2433, 1998.
- [187] A. Kleinschmidt, P. Sterzer, and G. Rees. Variability of perceptual multistability: from brain state to individual trait. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):988–1000, May 2012.
- [188] P. C. Klink, J. W. Brascamp, R. Blake, and R. J. van Wezel. Experience-driven plasticity in binocular vision. *Current Biology*, 20(16):1464–1469, 2010.

- [189] P. C. Klink, A. J. Noest, V. Holten, A. V. van den Berg, and R. J. van Wezel. Occlusion-related lateral connections stabilize kinetic depth stimuli through perceptual coupling. *Journal of Vision*, 9(10):20, 2009.
- [190] P. C. Klink, R. J. A. v. Wezel, and R. v. Ee. United we sense, divided we fail: context-driven perception of ambiguous visual stimuli. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):932–941, May 2012.
- [191] T. Knapen, J. Brascamp, J. Pearson, R. van Ee, and R. Blake. The role of frontal and parietal brain areas in bistable perception. *The Journal of Neuroscience*, 31(28):10293–10301, 2011.
- [192] S. Koelsch. Neural substrates of processing syntax and semantics in music. *Current opinion in neurobiology*, 15(2):207–212, 2005.
- [193] S. Koelsch. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, 46(1):179–190, 2009.
- [194] S. Koelsch, T. Fritz, K. Schulze, D. Alsop, and G. Schlaug. Adults and children processing music: an fMRI study. *Neuroimage*, 25(4):1068–1076, 2005.
- [195] S. Koelsch, T. Gunter, A. D. Friederici, and E. Schröger. Brain indices of music processing: “nonmusicians” are musical. *Journal of Cognitive Neuroscience*, 12(3):520–541, 2000.
- [196] S. Koelsch, S. Jentschke, D. Sammler, and D. Mietchen. Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44(3):476–490, 2007.
- [197] S. Koelsch, B.-h. Schmidt, and J. Kansok. Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, 39(5):657–663, 2002.

- [198] S. Koelsch, E. Schroger, and T. C. Gunter. Music matters: Preattentive musicality of the human brain. *Psychophysiology*, 39(01):38–48, 2002.
- [199] K. Koffka. *Principles of Gestalt Psychology*. Mimesis Edizioni, Place of publication not identified, Sept. 2014.
- [200] A. Kohn. Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*, 97(5):3155–3164, May 2007.
- [201] A. Kohn and J. A. Movshon. Neuronal Adaptation to Visual Motion in Area MT of the Macaque. *Neuron*, 39(4):681–691, 2003.
- [202] H. M. Kondo and M. Kashino. Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage*, 36(1):123–130, 2007.
- [203] H. M. Kondo and M. Kashino. Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *The Journal of Neuroscience*, 29(40):12695–12701, 2009.
- [204] H. M. Kondo, N. Kitagawa, M. S. Kitamura, A. Koizumi, M. Nomura, and M. Kashino. Separability and Commonality of Auditory and Visual Bistable Perception. *Cerebral Cortex*, page bhr266, Sept. 2011.
- [205] J. Kornmeier and M. Bach. Ambiguous Figures – What Happens in the Brain When Perception Changes But Not the Stimulus. *Frontiers in Human Neuroscience*, 6, Mar. 2012.
- [206] R. T. Krampe and K. A. Ericsson. Maintaining excellence: deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, 125(4):331, 1996.
- [207] N. Kraus and B. Chandrasekaran. Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8):599–605, 2010.

- [208] C. L. Krumhansl. Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17(1):53–80, 1995.
- [209] P. Kruse, H.-O. Carmesin, L. Pahlke, D. Strüber, and M. Stadler. Continuous phase transitions in the perception of multistable visual patterns. *Biological cybernetics*, 75(4):321–330, 1996.
- [210] P. Kruse and M. Stadler. Ambiguity in mind and nature. *Springer series in synergetics*, 64, 1995.
- [211] M. Kubovy, A. O. Holcombe, and J. Wagemans. On the lawfulness of grouping by proximity. *Cognitive psychology*, 35(1):71–98, 1998.
- [212] M. Kubovy and J. Wagemans. Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, 6(4):225–234, 1995.
- [213] M. Kubovy and M. Yu. Multistability, cross-modal binding and the additivity of conjoined grouping principles. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1591):954–964, Apr. 2012.
- [214] L. C. Lack. *Selective attention and the control of binocular rivalry*, volume 11. Mouton De Gruyter, 1978.
- [215] D. R. Ladd, R. Turnbull, C. Browne, C. Caldwell-Harris, L. Ganushchak, K. Swoboda, V. Woodfield, and D. Dediu. Patterns of individual differences in the perception of missing-fundamental tones. *Journal of Experimental Psychology: Human Perception and Performance*, 39(5):1386, 2013.
- [216] P. Ladefoged and D. E. Broadbent. Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1):98–104, 1957.
- [217] L. N. Law and M. Zentner. Assessing musical abilities objectively: construction and validation of the profile of music perception skills. 2012.

- [218] K. M. Lee, E. Skoe, N. Kraus, and R. Ashley. Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience*, 29(18):5832–5840, 2009.
- [219] S. R. Lehky. Binocular rivalry is not chaotic. *Proceedings of the Royal Society of London B: Biological Sciences*, 259(1354):71–76, 1995.
- [220] Leonard Bernstein. *The unanswered question: Six talks at Harvard*. Harvard University Press, 1976.
- [221] D. A. Leopold and N. K. Logothetis. Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, 3(7):254–264, July 1999.
- [222] D. A. Leopold, M. Wilke, A. Maier, and N. K. Logothetis. Stable perception of visually ambiguous patterns. *Nature neuroscience*, 5(6):605–609, 2002.
- [223] F. Lerdahl and R. Jackendoff. *A generative theory of tonal music*. MIT press, 1985.
- [224] D. J. Levitin. What does it mean to be musical? *Neuron*, 73(4):633–637, 2012.
- [225] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. Perception of the speech code. *Psychological review*, 74(6):431, 1967.
- [226] R. M. Liebert and B. Burk. Voluntary control of reversible figures. *Perceptual and motor skills*, 61(3f):1307–1310, 1985.
- [227] R. Lipp, P. Kitterick, Q. Summerfield, P. J. Bailey, and I. Paul-Jordanov. Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia*, 48(5):1417–1425, 2010.
- [228] G. M. Long and C. J. Moran. How to keep a reversible figure from reversing: teasing out top-down and bottom-up processes. *Perception*, 36(3):431, 2007.

- [229] G. M. Long and A. D. Olszewski. To reverse or not to reverse: When is an ambiguous figure not ambiguous? *The American journal of psychology*, 112(1):41, 1999.
- [230] G. M. Long and T. C. Toppino. Enduring Interest in Perceptual Ambiguity: Alternating Views of Reversible Figures. *Psychological Bulletin*, 130(5):748–768, 2004.
- [231] A. J. Lotto and K. R. Kluender. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60(4):602–619, 1998.
- [232] A. J. Lotto, S. C. Sullivan, and L. L. Holt. Central locus for nonspeech context effects on phonetic identification (L). *The Journal of the Acoustical Society of America*, 113(1):53–56, Jan. 2003.
- [233] R. D. Luce. *Response times: Their role in inferring elementary mental organization*. Number 8. Oxford University Press on Demand, 1986.
- [234] A. Maier, M. Wilke, N. K. Logothetis, and D. A. Leopold. Perception of Temporally Interleaved Ambiguous Patterns. *Current Biology*, 13(13):1076–1085, July 2003.
- [235] B. J. Malone, B. H. Scott, and M. N. Semple. Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. *The Journal of neuroscience*, 22(11):4625–4638, 2002.
- [236] L. T. Maloney, M. F. Dal Martello, C. Sahm, and L. Spillmann. Past trials influence perception of ambiguous motion quartets through pattern completion. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):3164–3169, 2005.
- [237] B. Maniscalco and H. Lau. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, 21(1):422–430, 2012.

- [238] B. Maniscalco and H. Lau. Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-dprime, Response-Specific Meta-dprime, and the Unequal Variance SDT Model. In *The cognitive neuroscience of metacognition*, pages 25–66. Springer, 2014.
- [239] V. A. Mann. Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28(5):407–412, 1980.
- [240] K. Masutomi and M. Kashino. Frequency-change aftereffect produced by adaptation to real and illusory unidirectional frequency sweeps. *The Journal of the Acoustical Society of America*, 134(1):EL14–EL18, 2013.
- [241] S. McAdams and J. Bertoncini. Organization and discrimination of repeating sound sequences by newborn infants. *The Journal of the Acoustical Society of America*, 102(5):2945–2953, Nov. 1997.
- [242] J. McDermott and M. Hauser. The origins of music: Innateness, uniqueness, and evolution. 2005.
- [243] J. H. McDermott, A. J. Lehr, and A. J. Oxenham. Individual differences reveal the basis of consonance. *Current Biology*, 20(11):1035–1041, 2010.
- [244] K. L. McDonald and C. Alain. Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *The Journal of the Acoustical Society of America*, 118(3):1593–1604, 2005.
- [245] J. P. Meenan and L. A. Miller. Perceptual flexibility after frontal or temporal lobectomy. *Neuropsychologia*, 32(9):1145–1149, 1994.
- [246] A. H. Mehta, I. Yasin, A. J. Oxenham, and S. Shamma. Neural correlates of attention and streaming in a perceptually multistable auditory illusion. *The Journal of the Acoustical Society of America*, 140(4):2225–2233, 2016.

BIBLIOGRAPHY

- [247] A. P. Merriam. *The anthropology of music*. Northwestern University Press, 1964.
- [248] L. B. Meyer. *Emotion and meaning in music*. University of Chicago Press, 2008.
- [249] C. Micheyl, K. Delhommeau, X. Perrot, and A. J. Oxenham. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing research*, 219(1):36–47, 2006.
- [250] S. M. Miller, B. D. Gynther, K. R. Heslop, G. B. Liu, P. B. Mitchell, T. T. Ngo, J. D. Pettigrew, and L. B. Geffen. Slow binocular rivalry in bipolar disorder. *Psychological medicine*, 33(04):683–692, 2003.
- [251] S. M. Miller, N. K. Hansell, T. T. Ngo, G. B. Liu, J. D. Pettigrew, N. G. Martin, and M. J. Wright. Genetic contribution to individual variation in binocular rivalry rate. *Proceedings of the National Academy of Sciences*, 107(6):2664–2668, 2010.
- [252] J. F. Mitchell, G. R. Stoner, and J. H. Reynolds. Object-based attention determines dominance in binocular rivalry. *Nature*, 429(6990):410–413, 2004.
- [253] M. Moerel, F. De Martino, R. Santoro, E. Yacoub, and E. Formisano. Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex. *NeuroImage*, 106:161–169, Feb. 2015.
- [254] B. C. Moore, B. R. Glasberg, and R. W. Peters. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, 80(2):479–483, 1986.
- [255] B. C. Moore and H. Gockel. Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88(3):320–333, 2002.
- [256] B. C. Moore, R. W. Peters, and B. R. Glasberg. Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America*, 77(5):1861–1867, 1985.

- [257] B. C. J. Moore and H. E. Gockel. Properties of auditory stream formation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1591):919–931, Apr. 2012.
- [258] S. Moreno, E. Bialystok, R. Barac, E. G. Schellenberg, N. J. Cepeda, and T. Chau. Short-term music training enhances verbal intelligence and executive function. *Psychological science*, 22(11):1425–1433, 2011.
- [259] S. Moreno and G. M. Bidelman. Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing research*, 308:84–97, 2014.
- [260] R. Moreno-Bote, A. Shpiro, J. Rinzel, and N. Rubin. Alternation rate in perceptual bistability is maximal at and symmetric around equidominance. *Journal of Vision*, 10(11):1, 2010.
- [261] G. Musacchia, M. Sams, E. Skoe, and N. Kraus. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40):15894–15898, 2007.
- [262] D. Müllensiefen, B. Gingras, J. Musil, L. Stewart, and others. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9(2):e89642, 2014.
- [263] W. Nager, C. Kohlmetz, E. Altenmüller, A. Rodriguez-Fornells, and T. F. Münte. The fate of sounds in conductors' brains: an ERP study. *Cognitive Brain Research*, 17(1):83–93, 2003.
- [264] T. Natsoulas. A study of the verbal-transformation effect. *The American journal of psychology*, pages 257–263, 1965.
- [265] L. A. Necker. Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. 1832.

- [266] D. L. Neff. Signal properties that reduce masking by simultaneous, random-frequency maskers. *The Journal of the Acoustical Society of America*, 98(4):1909–1920, 1995.
- [267] B. Nettl, N. L. Wallin, B. Merker, and S. Brown. An ethnomusicologist contemplates universals in musical sound and musical culture. *The origins of music*, pages 463–472, 2000.
- [268] A. J. Noest, R. Van Ee, M. M. Nijs, and R. J. A. Van Wezel. Percept-choice sequences driven by interrupted ambiguous stimuli: a low-level neural model. *Journal of vision*, 7(8):10, 2007.
- [269] M. S. Oechslin, D. Läge, and O. Vitouch. Training of tonal similarity ratings in non-musicians: a “rapid learning” approach. *Frontiers in psychology*, 3, 2012.
- [270] T. L. Ooi and Z. J. He. Binocular rivalry and visual awareness: the role of attention. *Perception-London*, 28(5):551–574, 1999.
- [271] J. Orbach, D. Ehrlich, and H. A. Heath. Reversibility of the Necker cube: I. An examination of the concept of "satiation of orientation". *Perceptual and motor skills*, 17(2):439–458, 1963.
- [272] A. J. Oxenham, B. J. Fligor, C. R. Mason, and G. Kidd Jr. Informational masking and musical training. *The Journal of the Acoustical Society of America*, 114(3):1543–1549, 2003.
- [273] A. Parbery-Clark, E. Skoe, and N. Kraus. Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, 29(45):14100–14107, 2009.
- [274] A. Parbery-Clark, E. Skoe, C. Lam, and N. Kraus. Musician enhancement for speech-in-noise. *Ear and hearing*, 30(6):653–661, 2009.
- [275] A. Parbery-Clark, D. L. Strait, S. Anderson, E. Hittner, and N. Kraus. Musical experience and the aging auditory system: implications for

- cognitive abilities and hearing speech in noise. *PLoS One*, 6(5):e18082, 2011.
- [276] C. V. Parise and C. Spence. 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, 4(5):e5664, 2009.
- [277] A. Pastukhov and J. Braun. A short-term memory of multi-stable perception. *Journal of vision*, 8(13):7–7, 2008.
- [278] A. D. Patel. *Music, language, and the brain*. Oxford university press, 2007.
- [279] A. D. Patel. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in psychology*, 2, 2011.
- [280] A. D. Patel. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing research*, 308:98–108, 2014.
- [281] A. D. Patel and J. R. Iversen. The linguistic benefits of musical abilities. *Trends in cognitive sciences*, 11(9):369–372, 2007.
- [282] A. D. Patel, J. R. Iversen, and J. C. Rosenberg. Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5):3034–3047, 2006.
- [283] J. Pearson, D. Tadin, and R. Blake. The effects of transcranial magnetic stimulation on visual rivalry. *Journal of Vision*, 7(7):2, 2007.
- [284] C. Pelofi, V. De Gardelle, P. Egger, and D. Pressnitzer. Inter-individual variability in auditory scene analysis revealed by confidence judgments (in press). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2016.

- [285] F. Perdreau and P. Cavanagh. Drawing skill is related to the efficiency of encoding object structure. *i-Perception*, 5(2):101–119, 2014.
- [286] A. Pinkus and A. Pantle. Probing Visual Motion Signals with a Priming Paradigm. *Vision Research*, 37(5):541–552, Mar. 1997.
- [287] C. J. Plack, A. J. Oxenham, and R. R. Fay. *Pitch: Neural Coding and Perception*. Springer Science & Business Media, Apr. 2006.
- [288] A. Platonov and J. Goossens. The role of lateral inhibition in binocular motion rivalry. *Journal of vision*, 13(6):12, 2013.
- [289] M. D. Plumbley. Efficient information transfer and anti-Hebbian neural networks. *Neural Networks*, 6(6):823–833, 1993.
- [290] A. Poellinger, R. Thomas, P. Lio, A. Lee, N. Makris, B. R. Rosen, and K. K. Kwong. Activation and habituation in olfaction—an fMRI study. *Neuroimage*, 13(4):547–560, 2001.
- [291] I. Pollack. Decoupling of auditory pitch and stimulus frequency: The Shepard demonstration revisited. *The Journal of the Acoustical Society of America*, 63(1):202–206, 1978.
- [292] A. Polonsky, R. Blake, J. Braun, and D. J. Heeger. Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature neuroscience*, 3(11):1153–1159, 2000.
- [293] B. Poulin-Charronnat, E. Bigand, and S. Koelsch. Processing of musical syntax tonic versus subdominant: an event-related potential study. *Journal of Cognitive Neuroscience*, 18(9):1545–1554, 2006.
- [294] B. Poulin-Charronnat, E. Bigand, F. Madurell, and R. Peereman. Musical structure modulates semantic priming in vocal music. *Cognition*, 94(3):B67–B78, 2005.
- [295] D. Pressnitzer and J.-M. Hupe. Temporal Dynamics of Auditory and Visual Bistability Reveal Common Principles of Perceptual Organization. *Current Biology*, 16(13):1351–1357, July 2006.

- [296] D. Pressnitzer, C. Suied, and S. A. Shamma. Auditory Scene Analysis: The Sweet Music of Ambiguity. *Frontiers in Human Neuroscience*, 5, Dec. 2011.
- [297] D. H. Raab. Backward masking. *Psychological Bulletin*, 60(2):118, 1963.
- [298] D. Raffman. Vagueness without paradox. *The Philosophical Review*, pages 41–74, 1994.
- [299] D. Raffman. *Unruly words: A study of vague language*. Oxford University Press, 2014.
- [300] V. S. Ramachandran and S. M. Anstis. Extrapolation of motion path in human visual perception. *Vision Research*, 23(1):83–85, 1983.
- [301] R. Ratcliff. Methods for dealing with reaction time outliers. *Psychological bulletin*, 114(3):510, 1993.
- [302] R. Ratcliff and J. N. Rouder. Modeling response times for two-choice decisions. *Psychological Science*, 9(5):347–356, 1998.
- [303] R. Ratcliff and J. J. Starns. Modeling confidence and response time in recognition memory. *Psychological review*, 116(1):59, 2009.
- [304] O. Raviv, M. Ahissar, and Y. Loewenstein. How recent history affects perception: the normative approach and its heuristic approximation. 2012.
- [305] J. E. Raymond and M. Isaak. Successive episodes produce direction contrast effects in motion perception. *Vision Research*, 38(4):579–589, 1998.
- [306] J. E. Raymond, H. L. O’Donnell, and S. P. Tipper. Priming reveals attentional modulation of human motion sensitivity. *Vision Research*, 38(19):2863–2867, 1998.

- [307] P. Regnault, E. Bigand, and M. Besson. Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials. *Journal of Cognitive Neuroscience*, 13(2):241–255, 2001.
- [308] B. H. Repp. The tritone paradox and the pitch range of the speaking voice: A dubious connection. *Music Perception*, pages 227–255, 1994.
- [309] B. H. Repp. Spectral envelope and context effects in the tritone paradox. *Perception*, 26(5):645–665, 1997.
- [310] B. H. Repp and G. Knoblich. Action can affect auditory perception. *Psychological Science*, 18(1):6–7, 2007.
- [311] B. H. Repp and J. M. Thompson. Context sensitivity and invariance in perception of octave-ambiguous tones. *Psychological research*, 74(5):437–456, 2010.
- [312] G. Rhodes, L. Jeffery, C. W. G. Clifford, and D. A. Leopold. The time-course of higher-level face aftereffects. *Vision Research*, 47(17):2291–2296, 2007.
- [313] C. Ricci and C. Blundo. Perception of ambiguous figures after focal brain lesions. *Neuropsychologia*, 28(11):1163–1173, 1990.
- [314] T. Rinne, M. H. Balk, S. Koistinen, T. Autti, K. Alho, and M. Sams. Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology*, 100(6):3323–3327, 2008.
- [315] C. Roberto and D. Jérôme. La philosophie du son. *Nîmes, Jacqueline Chambon*, 1994.
- [316] I. Rock, A. Gopnik, and S. Hall. Do young children reverse ambiguous figures? *Perception-London*, 23(6):635–644, 1994.
- [317] D. R. Ruggles, R. L. Freyman, and A. J. Oxenham. Influence of musical training on understanding voiced and whispered speech in noise. *PloS one*, 9(1), 2014.

- [318] D. R. Ruggles and A. J. Oxenham. Perceptual asymmetry induced by the auditory continuity illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3):908, 2014.
- [319] N. Russo, T. Nicol, G. Musacchia, and N. Kraus. Brainstem responses to speech syllables. *Clinical Neurophysiology*, 115(9):2021–2030, 2004.
- [320] J. R. Saffran, E. K. Johnson, R. N. Aslin, and E. L. Newport. Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1):27–52, 1999.
- [321] M. Sato, M. Baciú, H. Løevenbrück, J.-L. Schwartz, M.-A. Cathiard, C. Segebarth, and C. Abry. Multistable representation of speech forms: a functional MRI study of verbal transformations. *NeuroImage*, 23(3):1143–1151, 2004.
- [322] M. Sato, A. Basirat, and J.-L. Schwartz. Visual contribution to the multistable perception of speech. *Perception & psychophysics*, 69(8):1360–1372, 2007.
- [323] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29):8987–8992, 2015.
- [324] S. Schadwinkel and A. Gutschalk. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *Journal of neurophysiology*, 105(5):1977–1983, 2011.
- [325] E. G. Schellenberg, E. Bigand, B. Poulin-Charronnat, C. Garnier, and C. Stevens. Children’s implicit knowledge of harmony in Western music. *Developmental science*, 8(6):551–566, 2005.
- [326] G. Schlaug. The brain of musicians. *Annals of the New York Academy of Sciences*, 930(1):281–299, 2001.

- [327] G. Schlaug, L. Jancke, Y. Huang, and H. Steinmetz. In vivo evidence of structural brain asymmetry in musicians. *Science*, 267(5198):699–701, 1995.
- [328] G. Schlaug, L. Jäncke, Y. Huang, J. F. Staiger, and H. Steinmetz. Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8):1047–1055, 1995.
- [329] V. J. Schmithorst and M. Wilke. Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience letters*, 321(1):57–60, 2002.
- [330] M. A. Schmuckler. Expectation in music: Investigation of melodic and harmonic processes. *Music Perception: An Interdisciplinary Journal*, 7(2):109–149, 1989.
- [331] P. Schneider, V. Sluming, N. Roberts, M. Scherg, R. Goebel, H. J. Specht, H. G. Dosch, S. Bleeck, C. Stippich, and A. Rupp. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature neuroscience*, 8(9):1241–1247, 2005.
- [332] P. Schneider and M. Wengenroth. The neural basis of individual holistic and spectral sound perception. *Contemporary music review*, 28(3):315–328, 2009.
- [333] B. J. Scholl. Innateness and (Bayesian) visual perception. *The innate mind: Structure and contents*, page 34, 2005.
- [334] J.-L. Schwartz, N. Grimault, J.-M. Hupe, J.-M., B. C. J. Moore, and D. Pressnitzer. Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):896–905, May 2012.
- [335] O. Schwartz, A. Hsu, and P. Dayan. Space and time in visual context. *Nature Reviews Neuroscience*, 8(7):522–535, 2007.

- [336] C. M. Schwiedrzik, C. C. Ruff, A. Lazar, F. C. Leitner, W. Singer, and L. Melloni. Untangling perceptual memory: Hysteresis and adaptation map into separate cortical networks. *Cerebral cortex*, 24(5):1152–1164, 2014.
- [337] C. E. Seashore. *Manual of instructions and interpretations for measures of musical talent*. Columbia Graphophone Company, 1919.
- [338] A. Seither-Preisler, L. Johnson, K. Krumbholz, A. Nobbe, R. Patterson, S. Seither, and B. Lütkenhöner. Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3):743, 2007.
- [339] C. Semal and L. Demany. Individual differences in the sensitivity to pitch direction. *The Journal of the Acoustical Society of America*, 120(6):3907–3915, 2006.
- [340] M. Seppänen, E. Brattico, and M. Tervaniemi. Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning and Memory*, 87(2):236–247, 2007.
- [341] K. M. Shafritz, J. C. Gore, and R. Marois. The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences*, 99(16):10917–10922, 2002.
- [342] R. N. Shepard. Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, 36(12):2346–2353, 1964.
- [343] R. N. Shepard. Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4):305, 1982.
- [344] R. N. Shepard. Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, 1(1):2–28, 1994.
- [345] B. Shinn-Cunningham. Understanding informational masking from a neural perspective. *Proceedings of Meetings on Acoustics*, 19(1):060143, June 2013.

- [346] Z. J. Shu, N. V. Swindale, and M. S. Cynader. Spectral motion produces an auditory after-effect. 1993.
- [347] M. J. Sjerps, J. M. McQueen, and H. Mitterer. Evidence for pre-categorical extrinsic vowel normalization. *Attention, Perception, & Psychophysics*, 75(3):576–587, 2013.
- [348] J. Skerswetat, M. A. Formankiewicz, and S. J. Waugh. Very few exclusive percepts for contrast-modulated stimuli during binocular rivalry. *Vision Research*, 121:10–22, 2016.
- [349] J. A. Sloboda, J. W. Davidson, M. J. Howe, and D. G. Moore. The role of practice in the development of performing musicians. *British journal of psychology*, 87(2):287–309, 1996.
- [350] J. D. Smith, D. G. K. Nelson, L. A. Grohskopf, and T. Appleton. What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, 52(1):23–54, 1994.
- [351] J. S. Snyder, C. Alain, and T. W. Picton. Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation. *Journal of Cognitive Neuroscience*, 18(1):1–13, Jan. 2006.
- [352] J. S. Snyder, O. L. Carter, E. E. Hannon, and C. Alain. Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4):1232, 2009.
- [353] J. S. Snyder, O. L. Carter, S.-K. Lee, E. E. Hannon, and C. Alain. Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 34(4):1007, 2008.
- [354] J. S. Snyder, C. M. Schwiedrzik, A. D. Vitela, and L. Melloni. How previous experience shapes perception in different sensory modalities. *Frontiers in human neuroscience*, 9, 2015.

- [355] K. V. Sobel and R. Blake. How context influences predominance during binocular rivalry. *Perception*, 31(7):813, 2002.
- [356] M. F. Spiegel and C. S. Watson. Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, 76(6):1690–1695, 1984.
- [357] H. H. Spitz and R. S. Lipman. Some factors affecting Necker cube reversal rate. *Perceptual and Motor Skills*, 15(3):611–625, 1962.
- [358] N. Steinbeis and S. Koelsch. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5):1169–1178, 2008.
- [359] J. D. Stephens and L. L. Holt. Preceding phonetic context affects perception of nonspeech (L). *The Journal of the Acoustical Society of America*, 114(6):3036–3039, 2003.
- [360] P. Sterzer, A. Kleinschmidt, and G. Rees. The neural bases of multistable perception. *Trends in cognitive sciences*, 13(7):310–318, 2009.
- [361] P. Sterzer, M. O. Russ, C. Preibisch, and A. Kleinschmidt. Neural correlates of spontaneous direction reversals in ambiguous apparent visual motion. *Neuroimage*, 15(4):908–916, 2002.
- [362] C. E. Stilp, J. M. Alexander, M. Kiefte, and K. R. Kluender. Auditory color constancy: Calibration to reliable spectral properties across non-speech context and targets. *Attention, Perception, & Psychophysics*, 72(2):470–480, 2010.
- [363] D. L. Strait, N. Kraus, A. Parbery-Clark, and R. Ashley. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing research*, 261(1):22–29, 2010.
- [364] D. L. Strait, N. Kraus, E. Skoe, and R. Ashley. Musical experience and neural efficiency—effects of training on subcortical processing of vocal

- expressions of emotion. *European Journal of Neuroscience*, 29(3):661–668, 2009.
- [365] D. Strüber and C. S. Herrmann. MEG alpha activity decrease reflects destabilization of multistable percepts. *Cognitive Brain Research*, 14(3):370–382, 2002.
- [366] D. Strüber and M. Stadler. Differences in top down influences on the reversal rate of different categories of reversible figures. *Perception*, 28:1185–1196, 1999.
- [367] E. Sussman, R. Wong, J. Horváth, I. Winkler, and W. Wang. The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, 225(1–2):117–127, Mar. 2007.
- [368] J. Swaminathan, C. R. Mason, T. M. Streeter, V. Best, G. Kidd Jr, and A. D. Patel. Musical training, individual differences and the cocktail party problem. *Scientific reports*, 5, 2015.
- [369] O. Szalárdy, T. M. B\textbackslashtextbackslashHohm, A. Bendixen, and I. Winkler. Event-related potential correlates of sound organization: Early sensory and late cognitive effects. *Biological psychology*, 93(1):97–104, 2013.
- [370] S. Takei and S. Nishida. The effect of perceptual ambiguity in a bistable motion stimulus on perceptual latency. *The Journal of the Vision Society of Japan*, 21:151–163, 2009.
- [371] S. Takei and S. Nishida. Perceptual ambiguity of bistable visual stimuli causes no or little increase in perceptual latency. *Journal of vision*, 10(4):23, 2010.
- [372] C. Tallon-Baudry and O. Bertrand. Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, 3(4):151–162, 1999.

- [373] B. Tillmann, J. J. Bharucha, and E. Bigand. Implicit learning of tonality: a self-organizing approach. *Psychological review*, 107(4):885, 2000.
- [374] B. Tillmann, J. J. Bharucha, and E. Bigand. Implicit learning of regularities in Western tonal music by self-organization. In *Connectionist Models of Learning, Development and Evolution*, pages 175–184. Springer, 2001.
- [375] B. Tillmann and E. Bigand. Influence of global structure on musical target detection and recognition. *International Journal of Psychology*, 33(2):107–122, 1998.
- [376] B. Tillmann and E. Bigand. The relative importance of local and global structures in music perception. *The Journal of Aesthetics and Art Criticism*, 62(2):211–222, 2004.
- [377] B. Tillmann and B. Poulin-Charronnat. Auditory expectations for newly acquired structures. *The Quarterly Journal of Experimental Psychology*, 63(8):1646–1664, 2010.
- [378] D. Todorović. Context effects in visual perception and their explanations. *Review of Psychology*, 17(1):17–32, 2010.
- [379] F. Tong, M. Meng, and R. Blake. Neural bases of binocular rivalry. *Trends in cognitive sciences*, 10(11):502–511, 2006.
- [380] T. C. Toppino. Reversible-figure perception: Mechanisms of intentional control. *Perception & psychophysics*, 65(8):1285–1295, 2003.
- [381] S. E. Trehub. Toward a developmental psychology of music. *Annals of the New York Academy of Sciences*, 999(1):402–413, 2003.
- [382] N. Ulanovsky, L. Las, D. Farkas, and I. Nelken. Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*, 24(46):10440–10453, 2004.

- [383] L. P. A. S. van Noorden. *Temporal coherence in the perception of tone sequences*. Institute for Perceptual Research, 1975.
- [384] T. L. Van Zuijen, E. Sussman, I. Winkler, R. Näätänen, and M. Tervaniemi. Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. *Journal of cognitive neuroscience*, 16(2):331–338, 2004.
- [385] T. L. van Zuijen, E. Sussman, I. Winkler, R. Näätänen, and M. Tervaniemi. Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research*, 23(2):270–276, 2005.
- [386] J. A. Varela, K. Sen, J. Gibson, J. Fost, L. F. Abbott, and S. B. Nelson. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *The Journal of neuroscience*, 17(20):7926–7940, 1997.
- [387] D. Vickers and J. Packer. Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta psychologica*, 50(2):179–197, 1982.
- [388] E. K. Vogel and E. Awh. How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, pages 171–176, 2008.
- [389] M. W. von Grünau, S. Wiggin, and M. Reed. The local character of perspective organization. *Perception & Psychophysics*, 35(4):319–324, 1984.
- [390] M. Wallentin, A. H. Nielsen, M. Friis-Olivarius, C. Vuust, and P. Vuust. The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3):188–196, 2010.

- [391] M. Wang, D. Arteaga, and B. J. He. Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences*, 110(35):E3350–E3359, 2013.
- [392] N. Wang and A. J. Oxenham. Spectral motion contrast as a speech context effect. *The Journal of the Acoustical Society of America*, 136(3):1237–1245, 2014.
- [393] R. M. Warren. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *British Journal of Psychology*, 52(3):249–258, 1961.
- [394] R. M. Warren. Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, 70(4):261, 1968.
- [395] R. M. Warren and R. P. Warren. A comparison of speech perception in childhood, maturity, and old age by means of the verbal transformation effect. *Journal of Verbal Learning and Verbal Behavior*, 5(2):142–146, 1966.
- [396] D. Watanabe, T. Savion-Lemieux, and V. B. Penhune. The effect of early musical training on adult motor performance: evidence for a sensitive period in motor learning. *Experimental Brain Research*, 176(2):332–340, 2007.
- [397] M. A. Webster. Pattern selective adaptation in color and form perception. *The visual neurosciences*, 2:936–947, 2003.
- [398] R. L. Wegel and C. E. Lane. The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Physical Review*, 23(2):266–285, Feb. 1924.
- [399] A. Weise, E. Schröger, and A. Bendixen. The processing of concurrent sounds based on inharmonicity and asynchronous onsets: An object-related negativity (ORN) study. *Brain research*, 1439:73–81, 2012.

- [400] M. Wexler, M. Duyck, and P. Mamassian. Persistent states in vision break universality and time invariance. *Proceedings of the National Academy of Sciences*, 112(48):14990–14995, 2015.
- [401] K. L. Whiteford and A. J. Oxenham. Using individual differences to test the role of temporal and place cues in coding frequency modulation. *The Journal of the Acoustical Society of America*, 138(5):3093–3104, 2015.
- [402] F. A. Wichmann and N. J. Hill. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics*, 63(8):1293–1313, 2001.
- [403] M. A. Williams, A. P. Morris, F. McGlone, D. F. Abbott, and J. B. Mattingley. Amygdala responses to fearful and happy facial expressions under conditions of binocular suppression. *The Journal of Neuroscience*, 24(12):2898–2904, 2004.
- [404] H. R. Wilson. Binocular Rivalry: Neurons Unwire When They Can't Simultaneously Fire. *Current Biology*, 20(17):R715–R717, 2010.
- [405] I. Winkler, E. Kushnerenko, J. Horvath, R. Ceponiene, V. Fellman, M. Huotilainen, R. Näätänen, and E. Sussman. Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, 100(20):11812–11815, 2003.
- [406] L. Wittgenstein. *Philosophical investigations*. John Wiley & Sons, 2010.
- [407] P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, and N. Kraus. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*, 10(4):420–422, Apr. 2007.
- [408] K. Wunderlich, K. A. Schneider, and S. Kastner. Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature neuroscience*, 8(11):1595–1602, 2005.

- [409] R. Zatorre. Music, the food of neuroscience? *Nature*, 434(7031):312–315, 2005.
- [410] R. J. Zatorre, P. Belin, and V. B. Penhune. Structure and function of auditory cortex: music and speech. *Trends in cognitive sciences*, 6(1):37–46, 2002.
- [411] B. R. Zendel and C. Alain. Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21(8):1488–1498, 2009.
- [412] B. R. Zendel and C. Alain. Musicians experience less age-related decline in central auditory processing. *Psychology and aging*, 27(2):410, 2012.
- [413] B. R. Zendel and C. Alain. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of cognitive neuroscience*, 25(4):503–516, 2013.
- [414] W. Zhou and D. Chen. Binaral rivalry between the nostrils and in the cortex. *Current Biology*, 19(18):1561–1565, 2009.
- [415] W. Zhou, Y. Jiang, S. He, and D. Chen. Olfaction modulates visual perception in binocular rivalry. *Current Biology*, 20(15):1356–1358, 2010.
- [416] Y. H. Zhou, J. B. Gao, K. D. White, I. Merk, and K. Yao. Perceptual dominance time distributions in multistable visual perception. *Biological cybernetics*, 90(4):256–263, 2004.
- [417] B. J. Zimmerman and D. H. Schunk. An essential dimension of self-regulated learning. *Motivation and self-regulated learning: Theory, research, and applications*, page 1, 2008.

Appendix

Appendix A

Shepard tones and context effects

This poster presents the general method for the context effect of ambiguous intervals formed by two Shepard tones separated by a six semi-tone interval [342]. Here, two different types of prior information effects are reported. One is an hysteresis effect, by which the point of subjective indifference was modulated by the ordering presentation, one of the hallmarks of attractive context effects [63, 131]. The second reports the effect of a short sequence of tones played prior to the ambiguous pair. An attractive effect is observed: listeners tended to report the shift encompassing the frequency components of the context tones. The context effect accumulated with increasing number of tones in the sequence.

Perception of Ambiguous Auditory Stimuli: Hysteresis and Context Effects

Claire Chambers, Claire Pelofi, and Daniel Pressnitzer



Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes & Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France. Daniel.Pressnitzer@ens.fr



Overview

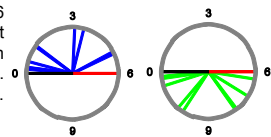
- Most investigations of pitch have used methods where the stimulus is manipulated and the effect on perception is recorded. Here we present a paradigm where context influences pitch perception in identical stimuli.
- Listeners judged the direction of pitch change in pairs of Shepard tones.
- In Expt. 1, we found strong assimilative hysteresis: the pitch change reported for one interval was influenced by previous percepts.
- In Expt. 2, we explored the causes of the hysteresis effect and found that it is likely to result from perceptual sensitisation.
- The paradigm may be of interest for investigating the neural bases of pitch, as activity can be recorded for different pitch percepts caused by the same stimulus.

Context Effect in Pitch Perception

Method

Stimuli

- Trial: ambiguous Shepard tone pair (6 st. interval), preceded by a context sequence of Shepard tones with random F0s ranging from either F0 to F0 + 6 st. (circle: upper path) or F0 to F0 - 6 st. (circle: lower path)



Procedure

- Task: Pitch direction (up/down) for the last two tones of each trial
- Participants: 8 normal-hearing listeners

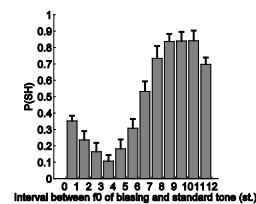
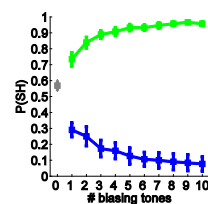
Results

Accumulation of bias

- Biases consistent with those found in Expt. 1
- Bias occurred when one context tone was presented
- Bias accumulated as the number of context tones was increased (F(10,60) = 121.20, p < 0.0001)

F0 of Shepard tones

- F0-interval of 3 st. and 9 st. relative to the standard tone resulted in the strongest biases, towards standard lower and standard higher percepts respectively (F(11, 77) = 19.02, p = 0.0001)

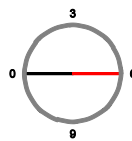
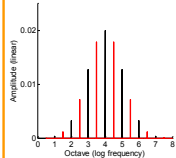


Hysteresis in Pitch Perception

Method

Stimuli

- Shepard tones (Shepard, 1964) with different fundamental frequencies (F0): Dominant cue for judging pitch direction is the proximity between components
- F0-interval of a half-octave (6 semitones, a tritone): Proximity cue is removed, resulting in ambiguous pitch shift (Shepard, 1964; Deutsch, 1986)



Sequence	Interval (semitones)
6 st.	6 6 6 6 6 6...
Random	3 6 8 4 1 11...
Increasing	1 2 3 4 5 6...
Decreasing	11 10 9 8 7 6...

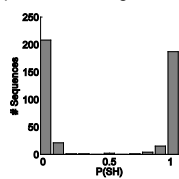
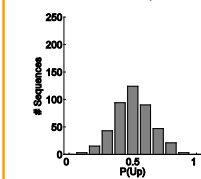
Procedure

- Shepard tone pairs were presented: Standard (black) and comparison (red)
- The F0 of the standard was varied across sequences
- Random order of standard and comparison
- Task: Pitch direction (up/down)
- Participants: 10 normal-hearing listeners

Results

Ambiguous 6 st. interval

- No systematic bias was found for "up" or "down" responses
- Standard or comparison systematically perceived as higher within series

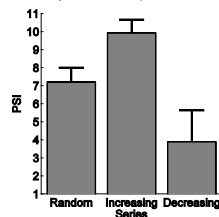
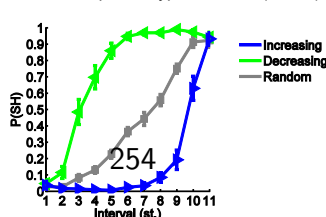


Random interval

- F0-intervals: 1 - 11 st. in random order

Ordered interval

- Hysteresis: 1 st. → 11 st., 11 st. → 1 st.
- Percept biased according to sequence type
- Point of subjective indifference (PSI) estimated by fitting a Weibull distribution to individual data
- Effect of sequence type on PSI: F(2, 192) = 282.42, p < 0.0001



Discussion

Expt. 1.

- Biases unlikely to derive from response hysteresis (Giangrande et al, 2003; Hock et al, 2005). Due to the random order within pairs, 'up' and 'down' responses were always balanced.
- Pitch was perceptually biased by the context.

Expt. 2.

- Both hysteresis (Expt. 1) and the context effect (Expt. 2) act like a form of perceptual sensitisation.
- Context acts in an implicit manner, outside the awareness of the listener
- Effect of context rapidly established. We suggest a role for rapid neural plasticity in producing the effect (Fritz et al, 2003).

Future Work

Future experiments will further characterise the context effect

- Specific to pitch stimuli or a more general phenomenon?
- The role of attention in producing the context effect
- The duration of the effect and the role played by attention in maintaining the effect (see companion poster, Pelofi, Chambers, & Pressnitzer)
- This paradigm provides a tool to investigate the neural bases of pitch perception, as neural activity can be recorded for different pitch percepts caused by the same stimulus, thus avoiding acoustical confounds

Acknowledgements and references

This work was supported by Entendre SAS and Fondation Pierre Gilles de Gennes.

- Deutsch (1987). The tritone paradox: effects of spectral variables. *Perception & Psychophysics*, 41, 563-575.
 Fritz, Shamma, Eihilali and Klein (2003) Rapid task-related plasticity of spectrottemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6, 1216-1223.
 Giangrande, Tuller, and Kelso (2003). Perceptual dynamics of circular pitch. *Music Perception*, 20, 241-262.
 Hock, Bukowski, Nichols, Huisman, and Rivera (2005). Dynamical vs. judgmental comparison: hysteresis effects in motion perception. *Spatial Vision*, 18, 317-335.
 Shepard, (1964). Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.*, 36, 2346-2353.

Appendix B

Effect of very short context tone

Expmt 1.3 revealed that a when a short silence was introduced before the test sequence, the build up of the context effect was even more pronounced. To further investigate that, n experiment was conducted that tested only very short context durations (10 and 20 ms) succeeded by silent gaps of various durations. The experiment design was:

Subj*Context2*Gap4

B.0.1 Method

Participants

Ten self-reported normal-hearing listeners ($M = 25.6$, $SD = 7.2$), four men and six women participated in the experiment. All of them successfully passed through the screening test (version 2, see [Screening procedure](#)). Among them, three had never participated in a previous experiment involving Shepard tones. All were paid for their contribution.

Stimuli

All the tones used in this experiment were Shepard tones generated as described above ([Stimuli](#)). Each trial consisted in the succession of a context sequence of one tone (C) and a test sequence composed of two test tones (T1-T2). The Fb for T1 was randomly drawn for each trial, uniformly between 60 Hz and 120 Hz. The interval between Fbs in T1 and T2 was fixed at 6 st which correspond to the ambiguous interval (for which listeners typically report two pitch-shift directions with equal chances). The interval between Fbs in C and T1 was fixed at 3 st and 9 st so that C would induce the maximal bias toward an upward (3 st) or a downward (9 st) pitch-shift. The duration of the test tones was 125ms each, including 2.5 ms raised-cosine onset and offset ramp. There was no delay between T1 and T2. The duration of C was varied between 10 and 20 ms. A silent gap was introduced between the context tone and the test tones. Its duration varied between 0, 0.0625, 0.125 and 0.25 s. Each gap and context duration combination was repeated forty times and the presentation order was shuffled within test sessions.

In order to minimize any across trial effect, an inter-trial sequence of five inter-trial tones was played, as previously described [Stimuli](#). Each inter-trial tone had a duration of 125 ms, including a 5ms raised-cosine onset and offset ramp and was separated from the subsequent tone by a short silence of 125 ms.

Apparatus and procedure

Listeners were tested individually in a double-walled sound-insulated booth (Industrial Acoustics Company). Stimuli were played diotically through an RME Fireface 800 soundcard, at a 16-bit resolution and a 44.1 kHz sampling rate. They were presented through Sennheiser HD 250 Linear II headphones. The presentation level was 65 dB SPL, A-weighted. At the end of each trial of the main experiment, listeners were asked to report whether the interval

formed by the two final tones was going upward or downward in pitch. To give their answer, they had to press either key 1 (downward shift) or key 2 (upward shift) of the keyboard. The inter-trial sequence was then played. During the control experiment, listeners had to report the pitch-shift direction, as in the main experiment. Then, they were asked to report whether they heard c tone before the final tone pair. To give their answer, they had to press either key 1 or key 2 of the keyboard. A green or red box appeared on the screen after each detection response to provide feedback. The inter-trial sequence was then played.

Data analysis

The proportion of "up" response $P(\text{up})$ was computed from the pitch-shift direction responses for all listeners and each gap duration condition. As previously described, a proportion of "biased" responses $P(\text{bias})$ was computed from the $P(\text{up})$. A $P(\text{bias})$ of 1 would correspond to listeners always reporting pitch shifts encompassing the frequency region of the context tones i.e. assimilative bias, whereas $P(\text{bias})$ of 0 would correspond to listeners always reporting the opposite direction of pitch shift, i.e. contrastive bias ; an absence of context effect, that is, a response probability unaffected by the context, would correspond to $P(\text{bias})$ of 0.5.

B.0.2 Results

Pitch-shift direction response

Figure B.0.1 displays the $P(\text{bias})$ averaged across participants is plotted as a function of the gap C-T1 duration for each context duration conditions in separated panels.

A two-way repeated measure ANOVA with gap and context durations as two within subject factors was conducted. It revealed that the gap factor had a

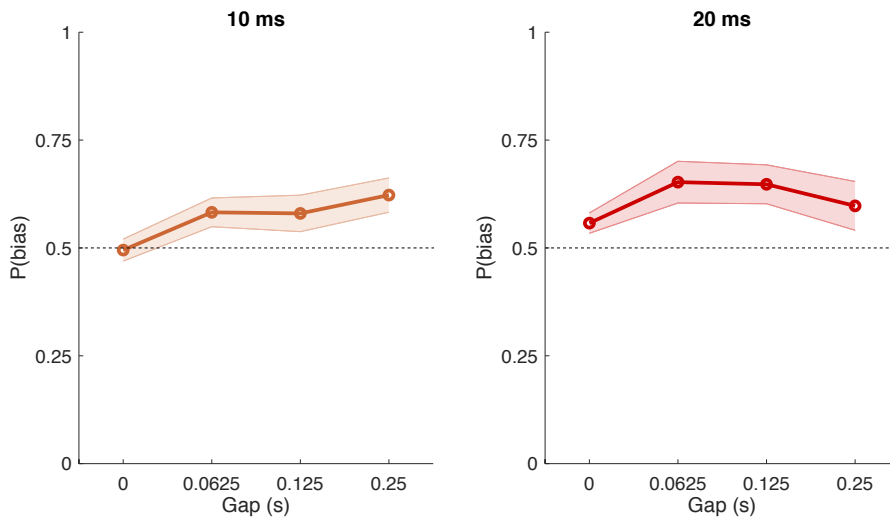


Figure B.0.1: The proportion of “biased” responses, $P(\text{bias})$, is displayed for each context duration and averaged across listeners. Shaded areas indicate ± 1 standard error about the mean.

significant effect ($F(3, 27) = 3.49, p = 0.03$) and the context duration factor barely reached significance ($F(1, 9) = 5.02, p = 0.051$). The interaction between the two was not significant ($F(3, 27) = 1.64, p = 0.2$).

B.0.3 Discussion

From the results of Expmt 1.3, we could have expected to observe a strong build up effect that would result from the interaction between short context and the silent gap preceding the test tone. The results presented here did confirm our expectations, as no strong build up was observed.

Appendix C

Can distraction tasks interrupt context effect in Shepard tone paradigm?

This poster presents preliminary results of a set of experiments exploring the effect of different types of distracting tasks on the disruption of a context effect. Chapter 5 of this manuscript reports remarkable remanence properties of the context effect of a short sequence of tones on an ambiguous interval [62]. Although the slow decay dynamics points to the contribution of non-sensory, higher-order cognitive processes [10, 286], as it has been argued with non-temporally adjacent context effects on speech perception [158] and visual motion perception [172], the exact nature of memory processes involved remain obscure. Especially, the strategies endorsed by participants during the silent gap are not controlled and could strongly modulate the remanence dynamics. By introducing a distraction task in between the context and the test sequence, we sought to control for listener's behavior and observe what would disrupt the context effect. The results presented in this poster are preliminary and would require further attention.

Appendix C Can distraction tasks interrupt context effect in Shepard tone paradigm?

Perception of ambiguous auditory stimuli: Memory and distractors

Claire Pelofi, Claire Chambers, and Daniel Pressnitzer



Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes & Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France. Daniel.Pressnitzer@ens.fr



Overview

We investigated context effects in the perception of ambiguous pitch stimuli, specifically addressing its duration and the role of attention in maintaining the context effect. The stimuli were pairs of "complementary" sounds, known as Shepard tones (Shepard, 1964).

In the companion poster (Chambers, Pelofi & Pressnitzer), it was found that it was possible to bias the direction of pitch shift by preceding the ambiguous Shepard tone pair with a context sequence of Shepard tones.

In Expt. 1, the duration of the effect was investigated by introducing a variable silent delay (0.5 s-64 s) between the context sequence and the ambiguous test. It was found that the effect lasted up to 32 seconds which suggests a rapidly established, long-lasting auditory context effect.

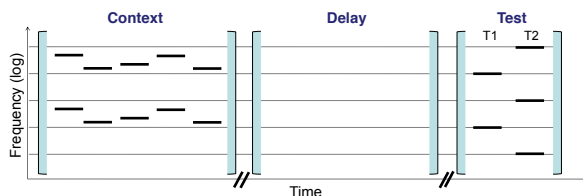
Expt. 2 examined the role of attention in maintaining the effect by requiring participants to respond to a distractor task during the silent delay. An auditory frequency discrimination and mental calculation task were used. Results suggest that the effect of attention is variable across participants.

Expt. 1: Implicit memory in the perception of ambiguous pitch stimuli

Method

Stimuli

- **Shepard tones** : tones which comprise all octaves of a given fundamental frequency. A Shepard tone pair with a half-octave interval is ambiguous in terms of the direction of pitch shift.
- **Context**: Five Shepard tones, with random F0s restricted to a half-octave range relative to T1.
- **Delay**: A silent gap of variable duration (0.5s-64s). Delays were varied randomly within a block.
- **Test**: A pair of Shepard tones, T1 and T2, with a half-octave F0-interval (Shepard, 1964). Random order of T1 and T2 on each trial.

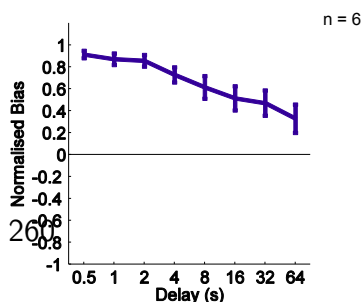


Procedure

- Participants indicated whether they heard upward or downward pitch change.
- Six normal-hearing listeners participated.

Results

- The proportion of trials where T1 was perceived as lower in pitch was computed. A normalised measure of bias was derived so that 0 indicates the absence of bias and 1 indicates full bias.
- Repeated measures ANOVA: significant decrease of the context effect over time ($F(7, 42) = 9.63, p < 0.001$).
- T-tests (Bonferroni correction): the normalised bias is significantly different from 0 for delay durations up to 32 s.

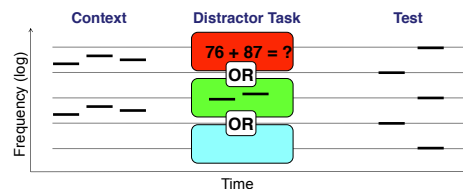


Expt. 2: Does attention maintain the bias?

Method

Stimuli

- As in Expt. 1 for context and test with delay(s): 8, 16, and 32. During the delay, distractor tasks were added. Three conditions were tested:
 - **Auditory frequency discrimination**: A component of the Shepard tones was randomly selected. Two pure tones were generated at frequencies of ± 0.25 st. relative to the selected component. Participants judged the direction of pitch-change.
 - **Mental calculation**: Participants added random two-digit numbers.
 - **Control**: Silent gap, as in Expt. 1.

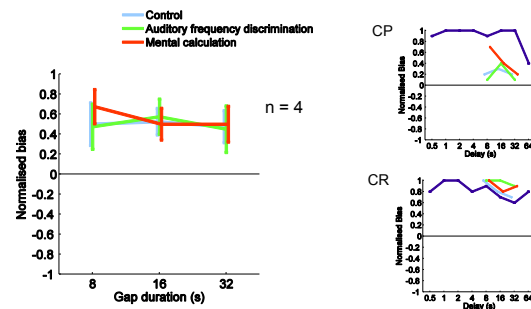


Procedure

- Following each trial, participants responded to the distractor task, then indicated the direction of change of the ambiguous tone pair.
- Six normal-hearing listeners participated, two were excluded.

Results

- Preliminary results reveal that there was no effect of distractor condition.
- However a comparison between Expt. 1 and 2 reveals two different patterns:
 - (i) a weaker context effect in Expt. 2 than Expt. 1 (participant CR)
 - (ii) an equally strong context effect in both experiments (participant CP)
- This dichotomy leads us to two hypotheses: (i) maintenance of the context effect requires more attention in certain listeners, (ii) the distracting task is more attentionally demanding for some participants.
- All participants scored more than 85% correct on all distractor tasks.



Rapid auditory context effect

Context effect in the perception of an ambiguous pitch stimulus

- Rapidly established (five context tones)
- Long lasting (persists up to 32s)
- Hypothesis (i) and (ii) to be tested in further research
- Possible behavioral correlate of rapid plasticity (Fritz et al, 2005)

Future research

- The present paradigm will be tested on amusic subjects (Supervisor: Isabelle Peretz, BRAMS laboratory, Montréal), who exhibit impaired music perception but intact peripheral auditory processing (Peretz et al., 2009).
- We will investigate the hypothesis that their impairment is linked to explicit processing involved in music perception and short memory for pitch information.

Acknowledgements and references

We would like to thank the Auditory Cognitive Neuroscience Erasmus Mundus Network for funding the follow-up to this work.
 J. Fritz, S. Shamma, M. Elhilali and D. Klein (2003) Rapid task-related plasticity of spectrotemporal receptive fields in Primary auditory cortex. *Nature Neuroscience*
 I. Peretz, E. Brattico, M. Jarvenpää and M. Tervaniemi. (2009) The amusic brain: in tune, out of key and unaware. *Brain*
 R.N. Shepard (1964). Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.*

Appendix D

Model of pre-perceptual grouping

This poster introduces a model proposed to account for the perception of an ambiguous interval of Shepard tones [342]. It is inspired by a factorial hidden Markov model [130] and claims that each frequency component in the acoustic scene is assigned to a perceptual track constrained by spectro-temporal continuity, in agreement with perceptual grouping principles at play in auditory scene analysis [47]. The properties of the perceptual tracks are continuously adapted, based on the integration of past information [156, 304, 166, 13].

Pre-perceptual grouping of auditory scenes explains contextual biases in the perception of ambiguous tonal shifts.



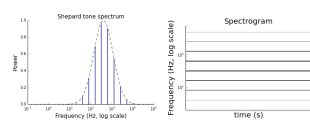
Vincent Adam¹, Claire Chambers², Claire Peloff², Daniel Pressnitzer², Maneesh Sahani¹

Gatsby Computational Neuroscience Unit, University College London
Ecole Normale Supérieure, Cognitive Studies Department

Background

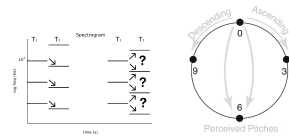
Shepard Tones

- Base frequency f_b , and octave related
- Bell-shaped envelope on log-frequency
- Evoke a non ambiguous pitch



Shift Ambiguity

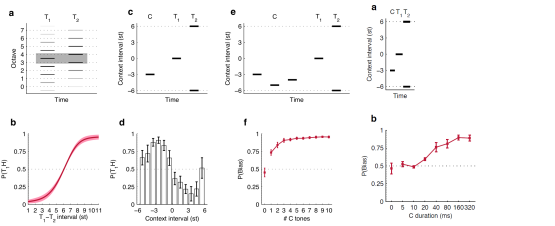
- Small Physical upward shift $f_b \uparrow$ leads to upward perceived shift
- Octave shift = Identity
- Half-Octave Shift \rightarrow Ambiguous percept



Biasing Pitch Shift Perception (Chambers and Pressnitzer) [3]

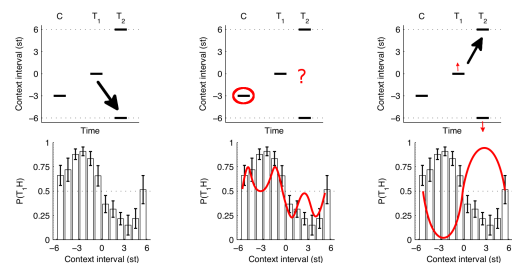
Psychophysical Results

- Demonstration of half-octave ambiguity. (a,b)
- A single Shepard tone played shortly prior to the ambiguous pair can induce a strong bias, toward the shift that encompasses that tone. (c,d)
- Build up of the bias with the number of preceding tones (e,f)

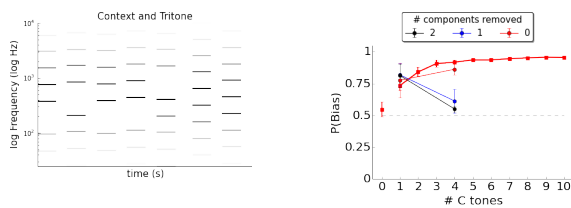


A simple adaptation mechanism?

- Adaptation is usually local
 - tonotopically local adaptation
 - effects falls off to 0 with distance
 - \rightarrow Here bias is strongest in the middle of the region
- Psychophysical manifestations of adaptation are usually repulsive
 - Englitz et al [2], showed repulsion in a decoding paradigm
 - repulsion predicts the opposite effect



- More than Average Spectrum of Context
 - Sparsely dropping individual pure tones removes the bias



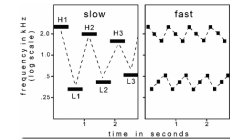
References

[1] R. Shepard, *JASA* 1964
[2] B. Englitz, et al *Basic Aspects of Learning* 2013
[3] C. Chambers, et al *submitted* 2015
[4] A. Bregman, *Auditory scene analysis* 1994
[5] Z. Ghahramani, Jordan, M *Machine learning* 1997

Model Concept: Preperceptual Grouping

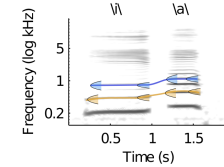
Inspiration: perceptual grouping in stream segregation [4].

- Items of an auditory scene can be segregated in multiple distinct perceptual objects



Preperceptual grouping

- Items of an auditory scene can be grouped in multiple distinct pre-perceptual tracks based on spectro-temporal continuity
- Helps solving the attribution problem (prevents misbinding)
- Assumption reflects natural sounds statistics



Model Presentation

Model assumptions

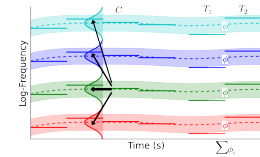
- sensory noise corrupt frequency representation
- track spectro-temporal continuity
 - tracks persist in time
 - tracks generate tones close to their mean (std σ_C)
 - sensory noise $\propto \frac{1}{\sqrt{\text{duration}}}$



Formalism: Factorial Hidden Markov Model [5]

Probabilistic inference:

- Online inference of the tracks underlying the auditory scene
- sensed tones are attributed to tracks (clustered)
- beliefs on tracks center frequency are updated



Variational Approximation: $p(T_i^{(i)}|H_i) \approx \mathcal{N}(T_i^{(i)}|\mu_i^{(i)}, \sigma_i^{(i)})$

Tone Attribution:

$$r_k^{(i)}(Y) = p(r_k^{(i)} = k|Y, H)$$

Tracks update:

$$\mu^{(i)} \leftarrow \left(\frac{\mu^{(i)}}{\sigma^{(i)2}} + \frac{\sum_{k=0}^L r_k^{(i)} c_k}{\sigma_C^2} \right) / \left(\frac{1}{\sigma^{(i)2}} + \frac{\sum_{k=0}^L r_k^{(i)}}{\sigma_C^2} \right)$$

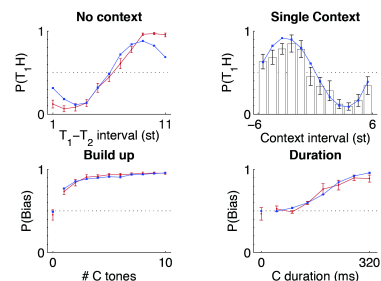
$$\sigma^{(i)2} \leftarrow \left(\frac{1}{\sigma^{(i)2}} + \frac{\sum_{k=0}^L r_k^{(i)}}{\sigma_C^2} \right)^{-1}$$

From pre-perceptual tracks to perception:

Track-local features are extracted and combined to give rise to a percept. Here, track-local shifts are linearly combined to give rise to the global shift percept.

Model Fitting Results

The model predictions qualitatively and quantitatively match the psychophysical results. It explains perceived shift both in the presence and the absence of context.



Discussion

Summary

- Chambers et al. presented a contextual bias in auditory perception
- We suggest that a simple adaptive mechanism cannot account for the bias
- We propose a model relying on pre-perceptual tracking of auditory scenes to account for the bias and show it matches psychophysical results

Pre-perceptual Grouping, a general mechanism?

Résumé

La tâche des systèmes sensoriels est d'organiser l'information lacunaire et ambiguë en une représentation stable et cohérente du monde. Dans la modalité auditive, « l'analyse de scènes auditives » consiste à regrouper ou séparer correctement le signal en différentes voix.

Cette thèse a été dédiée à caractériser la manière dont est perçu un stimulus auditif ambigu. Le stimulus est composé de deux sons qui forment un intervalle qui peut être entendu comme montant ou descendant en hauteur.

Dans un premier temps, mon but a été de mettre en évidence le décours temporel de l'effet de contexte permettant d'orienter la perception ambiguë dans un sens ou dans l'autre. Les dynamiques temporelles remarquables de cet effet de contexte ont permis de déterminer plus précisément la nature des mécanismes neuronaux sous-jacents.

Dans un deuxième temps, ce travail de recherche s'est penché sur une question nouvelle : l'expérience subjective du stimulus ambigu. Grâce à la combinaison de mesures de confiance et de temps de réponses, les résultats ont mis en évidence un effet prononcé et robuste de l'expertise musicale.

Cette thèse présente donc une nouvelle approche de la perception d'un stimulus auditif ambigu qui permet d'aborder la question encore largement inexplorée de l'expérience subjective de l'ambiguïté.

Mots clés

Ambiguïté, Hauteur, Effet de Contexte, Expertise Musicale

Abstract

Perceptual systems are endorsed with the challenging task to organize a partial and ambiguous sensory input into a useful representation of the world. In the auditory modality this task named "auditory scene analysis" consists in grouping or streaming the acoustic input into different voices.

The aim of this Ph.D has been to investigate how listeners process an ambiguous auditory stimulus. The stimulus consists in two tones that can be heard as an upward or a downward pitch-shift.

First, I investigated the sensitivity of the ambiguous interval to prior contextual information and the time-course of these context effects. The remarkable temporal dynamics of the context effect are informative of the underlying neural processes at play.

Second, my research tackled a novel question on perceptual ambiguity: are listeners aware of the existence of the distinct perceptual alternatives? By combining confidence ratings and response times, the results revealed a strong and robust effect of musical expertise.

This work proposes a novel approach to an ambiguous auditory stimulus, to tackle the largely unexplored question of the subjective experience of ambiguity.

Keywords

Ambiguity, Pitch, Context Effects, Musicianship