



Complex-tone pitch representations in the human auditory system.

Bianchi, Federica

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Bianchi, F. (2016). *Complex-tone pitch representations in the human auditory system*. Technical University of Denmark, Department of Electrical Engineering. Contributions to hearing research Vol. 22

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

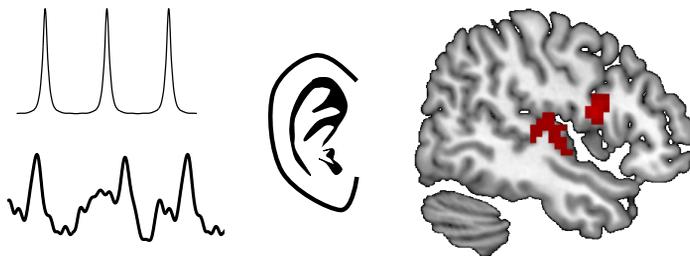
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 22

Federica Bianchi

Complex-tone pitch representations in the human auditory system



Complex-tone pitch representations in the human auditory system

PhD thesis by
Federica Bianchi



Technical University of Denmark

2016

© Federica Bianchi, 2016
The defense was held on June 10, 2016.

This PhD dissertation is the result of a research project carried out at the Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark (Kgs. Lyngby, Denmark) and at the Danish Research Center for Magnetic Resonance Imaging (DRCMR, Hvidovre Hospital, Hvidovre, Denmark). Part of the project was carried out at the Montreal Neurological Institute (MNI, Montreal, Canada) and at the International laboratory for Brain, Music, and Sound Research (BRAMS, Montreal, Canada).

The project was financed by the Technical University of Denmark. The external stay in Montreal was further supported by a scholarship from the Erasmus Mundus in Auditory Cognitive Neuroscience. All conferences and additional visits in Montreal were financed by the EliteForsk-rejsestipendium grant.

Supervisors

Main supervisor

Prof. Torsten Dau

Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Co-supervisors

Ass. Prof. Sébastien Santurette

Jens Hjortkjaer, PhD

Dorothea Wendt, PhD

Hearing Systems Group
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

External advisors

Prof. Robert Zatorre

Montreal Neurological Institute
McGill University and BRAMS
Montreal, QC, Canada

Prof. Hartwig Siebner

Danish Research Centre for Magnetic Resonance (DRCMR)
Hvidovre Hospital
Hvidovre, Denmark

"Ora tu pensa: un pianoforte. I tasti iniziano. I tasti finiscono. Tu sai che sono 88, su questo nessuno può fregarti. Non sono infiniti, loro. Tu, sei infinito, e dentro quei tasti, infinita è la musica che puoi fare. Loro sono 88. Tu sei infinito."

"Take a piano. The keys begin. The keys end. You know that they are 88 and no one can tell you differently. They are not infinite, they. You, you are infinite, and within those keys, infinite is the music that you can make. They are 88. You are infinite."

Novecento (Alessandro Baricco)

Abstract

Understanding how the human auditory system processes the physical properties of an acoustical stimulus to give rise to a pitch percept is a fascinating aspect of hearing research. Since most natural sounds are harmonic complex tones, this work focused on the nature of pitch-relevant cues that are necessary for the auditory system to retrieve the pitch of complex sounds. The existence of different pitch-coding mechanisms for low-numbered (spectrally resolved) and high-numbered (unresolved) harmonics was investigated by comparing pitch-discrimination performance across different cohorts of listeners, specifically those showing enhanced pitch cues (i.e., musicians) and those typically having disrupted pitch cues (i.e., hearing-impaired listeners). In particular, two main topics were addressed: the relative importance of resolved and unresolved harmonics for normal-hearing (NH) and hearing-impaired (HI) listeners and the effect of musical training for pitch discrimination of complex tones with resolved and unresolved harmonics.

Concerning the first topic, behavioral and modeling results in listeners with sensorineural hearing loss (SNHL) indicated that temporal envelope cues of complex tones with unresolved harmonics may be enhanced relative to NH listeners at the output of peripheral auditory filters. This enhancement of temporal envelope coding was found to be ascribed to a reduction of cochlear compression. Since frequency selectivity and temporal fine structure (TFS) cues are known to be degraded in listeners with SNHL, it is likely that HI listeners rely on the enhanced envelope cues to retrieve the pitch of unresolved harmonics. Hence, the relative importance of pitch cues may be altered in HI listeners, whereby envelope cues may be used instead of TFS cues to obtain a similar performance in pitch discrimination to that of NH listeners.

In the second part of this work, behavioral and objective measures of pitch discrimination were carried out in musicians and non-musicians. Musicians showed an increased pitch-discrimination performance relative to non-musicians for both resolved and unresolved harmonics, although their benefit was larger for the resolved harmonics. Additionally, task-evoked pupil responses were recorded as an indicator of processing effort while listeners performed a pitch-discrimination task.

Although the difficulty of the task was adjusted for each participant to compensate for the individual pitch-discrimination abilities, the musically trained listeners still allocated lower processing effort than did the non-musicians to perform the task at the same performance level. This finding suggests an enhanced pitch representation along the auditory system in musicians, possibly as a result of training, which seemed to be specific to the stimuli containing resolved harmonics.

Finally, a functional magnetic resonance imaging paradigm was used to examine the response of the auditory cortex to resolved and unresolved harmonics in musicians and non-musicians. The neural responses in musicians were enhanced relative to the non-musicians for both resolved and unresolved harmonics in the right auditory cortex, right frontal regions and inferior colliculus. However, the increase in neural activation in the right auditory cortex of musicians was predictive of the increased pitch-discrimination performance only for resolved harmonics. These results suggest a training-dependent effect in musicians that is partially specific to the resolved harmonics.

Resumé

Forståelse af hvordan menneskets auditive system behandler de fysiske egenskaber af et akustisk stimulus, så det giver anledning til en opfattelse af tonehøjde, er et fascinerende aspekt af høreforskningen. Da de fleste naturligt forekommende lyde er harmoniske tonekomplekser, fokuserer denne afhandling på karakteren af de signaler, der er nødvendige, når det auditive system skal fastlægge tonehøjden af komplekse lyde. Eksistensen af forskellige mekanismer for kodning af tonehøjde for henholdsvis lavfrekvente (spektralt opløste) og højfrekvente (spektralt uopløste) overtoner er undersøgt ved at sammenligne diskriminationsevnen på tværs af forskellige lyttergrupper, særligt lyttere med gode diskriminationsevner (musikere) og lyttere med nedsatte evner (hørehæmmede lyttere). To hovedemner er behandlet: (1) den relative betydning af opløste og uopløste overtoner for normalthørende og hørehæmmede lyttere; og (2) effekten af musikalsk træning for diskrimination af tonehøjden for opløste og uopløste overtoner.

Under første hovedemne indikerer resultaterne, at detekteringen af den temporale indhylningskurve for komplekse toner er bedre hos lyttere med sensorineuralt høretab (SNHL) end hos normalthørende. Denne forbedrede detektering af den temporale indhylningskurve kan formentlig tilskrives en reduceret kompression i det indre øre hos de hørehæmmede. Da frekvensselektivitet og detektering af temporal finstruktur (TFS) er reduceret hos lyttere med SNHL, er det sandsynligt, at hørehæmmede lyttere udnytter deres bedre detektering af indhylningskurven til at afkode tonehøjden af uopløste overtoner. Formentlig anvender hørehæmmede i højere grad end normalthørende indhylningskurven i stedet for TFS til bestemmelse af tonehøjden. Dermed kan de opnå en tonehøjde-diskrimination på niveau med normalthørende lyttere.

I anden del af afhandlingen er adfærdsmæssige og objektive målinger af tonehøjde-diskrimination foretaget for musikere og ikke-musikere. Musikere udviste en øget tonehøjde-diskrimination i forhold til ikke-musikere for både opløste og uopløste overtoner, dog mest markant for de opløste overtoner. Derudover blev lytternes pupilrespons registreret som en indikator for bearbejdningsindsats under diskriminationsopgaven. Selvom opgavens vanskelighed blev justeret som kompensation for individuelle forskelle i diskriminationsevne, kunne de musikalsk trænedes lyttere

nøjes med en mindre indsats i forhold til ikke-musikere for at udføre opgaven. Dette tyder på en forbedret kodning af tonehøjden i det auditive system hos musikere, især for opløste overtoner. Endelig blev funktionel magnetisk resonansbilleddannelse (fMRI) anvendt til at undersøge responset i auditiv cortex for både opløste og uopløste overtoner hos musikere og ikke-musikere. De neurale reaktioner i flere dele af hjernen var forøget hos musikerne i forhold til ikke-musikerne for begge typer af overtoner. For de opløste overtoner var aktivitetsstigningen i højre auditive cortex i overensstemmelse med en øget tonehøjde-diskrimination for disse toner hos musikere. Disse resultater tyder på en træningsafhængig effekt, der er delvist specifik for de opløste overtoner.

Acknowledgments

A very special thanks goes to my supervisors, who have been of great support during the last three years. In particular, thank you, Torsten, for believing in me before I could actually believe in myself. Thank you for giving me the chance of doing a PhD, for giving me the freedom to explore different aspects of hearing science, and for giving me the possibility to travel so much during these years. Thank you for trusting in me, for pushing me to give presentations and stand for challenges. It was an amazing journey, every day. A journey through science and a journey through myself.

Thank you, Sébastien, for being always there to discuss the intriguing mechanisms of pitch perception. Thank you, Jens, for all the support during the fMRI and the data analysis. Thank you, Doro, for the nice collaboration on the pupillometry project and for supporting me during some sporadic crises during my PhD. Thank you, Sébastien, Doro and Jens for the great support during the last days before my hand in. A big thanks goes also to Robert Zatorre, for introducing me to the science of music, and for being always available for fascinating discussions.

Thank you to my office mates for standing me (some days more and some days less!) for three long years. Thank you for making our office the cosiest place to work. And thank you to all the colleagues and friends in the Hearing Systems group that brought a lot of fun and happiness in my life.

A big, big, big thanks goes to my family, to my little sister, and to all the amazing friends around the world. Thank you for your unconditional love, support and trust.

Federica Bianchi, April 20, 2016

Related publications

Journal papers

- Bianchi, F, Verhulst, S., Dau, T. (2013). “Experimental Evidence for a Cochlear Source of the Precedence Effect,” *J. Assoc. Res. Otolaryngol.* **14**, 767-779.
- Verhulst, S., Bianchi, F, Dau, T. (2013). “Cochlear contributions to the precedence effect,” *Advances in Experimental Medicine and Biology* **787**, 283-291.
- Bianchi, F, Santurette, S., Wendt, D., Dau, T. (2016). “Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort,” *J. Assoc. Res. Otolaryngol.* **17**, 69-79.
- Bianchi, F, Fereczkowski, M., Zaar, J., Santurette, S., Dau, T. (in press). “Complex-tone pitch discrimination in listeners with sensorineural hearing loss,” *Trends in Hearing*.
- Bianchi, F, Hjortkjær, J., Santurette, S., Siebner, H., Zatorre, R., Dau, T. (in prep). “Cortical correlates of complex-tone pitch discrimination in musicians and non-musicians,” *J. Neurosci*.

Conference papers

- Bianchi, F, Santurette, S., Wendt, D., Dau, T. (2014). “Objective correlates of pitch salience using pupillometry,” *Proc. of Forum Acusticum*, Krakow,

Poland, September 2014.

- Pelzer, A., Santurette, S., Bianchi, E., Dau, T. (2015). “Fast assessment of auditory spectral and temporal resolution,” *Proc. of Jahrestagung für Akustik (DAGA)*, pp. 932-935, Nuremberg, Germany, March 2015.
- Bianchi, E., Fereczkowski, M., Zaar, J., Santurette, S., Dau, T. (2016). “Effects of cochlear compression and frequency selectivity on pitch discrimination of complex tones with unresolved harmonics,” *Proc. of International Symposium on Auditory and Audiological Research (ISAAR): Individual hearing loss - Characterization, modelling, compensation strategies*, pp. 51-58, Nyborg, Denmark, August 2015.
- Santurette, S., Bianchi, E., Fereczkowski, M., Zaar, J., Dau, T. (2016). “Effets d’une compression cochléaire et d’une sélectivité en fréquence réduites sur la discrimination de la hauteur de tons complexes,” *Proc. of Proceedings of the French Acoustics Congress*, pp. 805-810, Le Mans, France, April 2016.

Published abstracts

- Bianchi, E., Verhulst, S., Dau, T. (2012). “Behavioral and objective measures of the precedence effect,” *Association for Research in Otolaryngology (ARO), 35th Mid-Winter Meeting*, San Diego, CA, February 2012.
- Bianchi, E., Santurette, S., Dau, T., Hjortkjær, J., Siebner, H. (2013). “Neural correlates of pitch salience using fMRI,” *The 7th Arches Meeting*, Paris, France, November 2013.
- Bianchi, E., Santurette, S., Dau, T. (2014). “Pitch coding of complex tones in the normal and impaired auditory system,” *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe, CA, August 2014.
- Bianchi, E., Santurette, S., Wendt, D., Dau, T. (2015). “Effects of musical training on pitch discrimination of resolved and unresolved complex tones,”

Association for Research in Otolaryngology (ARO), 38th Mid-Winter Meeting, Baltimore, MD, February 2015.

- Bianchi, F, Fereczkowski, M., Santurette, S., Dau, T. (2015). “Relation between temporal envelope coding, pitch discrimination, and compression estimates in listeners with sensorineural hearing loss,” *Acoustical Society of America (ASA), 167th Meeting, Pittsburgh, PA, May 2015.*
- Bianchi, F, Hjortkjær, J., Santurette, S., Siebner, H., Zatorre, R., Dau, T. (2015). “Cortical pitch representations of complex tones in musicians and non-musicians,” *International laboratory for Brain, Music, and Sound Research (BRAMS), 10th Anniversary Symposium, Montreal, Canada, October 2015.*
- Bianchi, F, Hjortkjær, J., Santurette, S., Siebner, H., Zatorre, R., Dau, T. (2016). “Complex-tones cortical pitch representations in musicians vs. non-musicians,” *Association for Research in Otolaryngology (ARO), 39th Mid-Winter Meeting, San Diego, CA, February 2016.*

Contents

Abstract	vii
Resumé på dansk	ix
Acknowledgments	xi
Related publications	xiii
Table of contents	xxi
1 Introduction	1
1.1 What is pitch and why is pitch important?	1
1.2 Complex tones and harmonic resolvability	2
1.3 Complex-tone pitch coding mechanisms	4
1.4 Pitch perception and hearing loss	6
1.5 Pitch perception and musical experience	7
1.6 Overview of the thesis	8
2 Complex-tone pitch discrimination in listeners with sensorineural hearing loss	11
2.1 Introduction	12
2.2 Methods	15
2.2.1 Listeners and experimental setup	15
2.2.2 Experiment I: Pitch-discrimination of complex tones	15
2.2.3 Experiment II: Amplitude-modulation detection	17

2.2.4	Experiment III: Estimates of BM I/O function and cochlear compression	18
2.2.5	Modeling the effects of cochlear compression and frequency selectivity on envelope peakiness	19
2.3	Results	21
2.3.1	Experiment I: Pitch-discrimination of complex tones	21
2.3.2	Experiment II: Amplitude-modulation detection	23
2.3.3	Experiment III: Estimates of BM I/O function and cochlear compression	24
2.3.4	Effects of cochlear compression and frequency selectivity on pitch discrimination	26
2.3.5	Modeling the effects of cochlear compression and frequency selectivity on envelope peakiness	28
2.4	Discussion	30
2.4.1	Relation between behavioral results and envelope representation	30
2.4.2	F_0 DL ratio and individual measures of cochlear compression and filter bandwidth	33
2.4.3	Modeling results and envelope enhancement	34
2.4.4	AM-detection as a measure of frequency selectivity	36
2.5	Conclusion	36
3	Objective correlates of pitch salience using pupillometry	39
3.1	Introduction	40
3.2	Experiment I: Difference limens for fundamental frequency	42
3.2.1	Method	42
3.2.2	Results	44
3.3	Experiment II: Pupillometry	45
3.3.1	Method	45
3.3.2	Results	48
3.4	Discussion	50
3.5	Conclusion	53

4 Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort	55
4.1 Introduction	56
4.2 Method	58
4.2.1 Experiment 1: Behavioral Pitch Discrimination Thresholds	58
4.2.2 Experiment 2: Effects of Harmonic Resolvability and Task Difficulty	60
4.2.3 Experiment 3: Pupil Dilations During Pitch Discrimination	62
4.3 Results	64
4.3.1 Behavioral Pitch Discrimination Thresholds	64
4.3.2 Effects of Harmonic Resolvability and Task Difficulty	66
4.3.3 Pupil Dilations During Pitch Discrimination	66
4.4 Discussion	72
4.5 Conclusion	77
5 Cortical correlates of complex-tone pitch discrimination in musicians and non-musicians	79
5.1 Introduction	80
5.2 Method	83
5.2.1 Listeners	83
5.2.2 Experiment I: Behavioral pitch-discrimination of complex tones	83
5.2.3 Experiment II: Cortical responses during a pitch-discrimination task	85
5.3 Results	89
5.3.1 Experiment I: Pitch-discrimination of complex tones	89
5.3.2 Experiment II: Cortical responses during a pitch-discrimination task	92
5.4 Discussion	96
5.4.1 The musicians' benefit in pitch discrimination for resolved and unresolved harmonics	96

5.4.2	The musicians' benefit and implications for a post-peripheral advantage	99
5.4.3	Neural correlates of resolvability	100
5.4.4	Correlates of musical training in the right primary auditory cortex	102
5.5	Conclusion	104
6	General discussion	107
6.1	Summary and discussion of main results	107
6.1.1	Pitch discrimination performance of hearing-impaired listeners: Importance of envelope cues	107
6.1.2	Pitch discrimination performance of musicians vs non-musicians: Effort and plasticity	109
6.2	Implications for pitch coding mechanisms	113
6.3	Perspectives	115
	Bibliography	117
A	Experimental Evidence for a Cochlear Source of the Precedence Effect	131
A.1	Introduction	132
A.2	Methods	134
A.2.1	Psychoacoustical experiments	134
A.2.2	Statistical analysis	139
A.3	Results	143
A.3.1	Psychoacoustical experiments	143
A.3.2	CEOAE	145
A.3.3	ABR	147
A.4	Discussion	149
A.4.1	Effect of frequency range and implications for peripheral processing	149
A.4.2	Effects of peripheral processing on the PE	152
A.4.3	CEOAEs and monaural ABRs	155
A.4.4	Contributions of binaural processes	155

A.5 Conclusion	156
A.6 References	157
Collection volumes	165

1

General introduction

1.1 What is pitch and why is pitch important?

Pitch is a subjective attribute of hearing which allows us to hear sounds as musical (Santurette, 2011). It plays a central role in our everyday experience, where it is used as a cue in many perceptual contexts. In particular, pitch is essential for music perception, allowing us to hear melodies and chords. Pitch is also important for speech recognition, carrying prosody information in most languages, as well as semantic information in tone languages. Additionally, pitch facilitates the perceptual segregation of sound sources, for example a female from a male talker, and can also be used for grouping individual sound frequency components in a unique pitch percept, for example the note of a piano (Plack et al., 2005).

Understanding how the human auditory system processes the physical properties of an acoustical stimulus (e.g., frequency content, repetition rate, modulation depth) to give rise to a musical percept is a fascinating aspect of fundamental hearing research. Most natural pitch-evoking sounds are harmonic complex tones, for example voiced phonemes in speech, any musical sound produced by a tonal musical instrument, as well as sounds produced by sound sources with a characteristic rate of vibration. Thus, clarifying the nature of pitch-relevant information that the auditory system uses to retrieve the pitch of a complex sound is of central importance towards understanding our abilities in recognizing, differentiating and eventually enjoying natural acoustic scenes.

Additionally, since it is well established that hearing loss affects the ability of listeners to perceive pitch, understanding the basic mechanisms underlying pitch perception in the normal auditory system is an essential step towards restoring

accurate pitch perception in the hearing-impaired population.

1.2 Complex tones and harmonic resolvability

Since most natural sounds are harmonic complex tones, understanding how the human auditory system retrieves the pitch of such complex sounds remains a crucial aspect of hearing research. A harmonic complex tone consists of a series of harmonic components at frequencies corresponding to integer multiples of the fundamental frequency (F_0) that typically elicit a pitch corresponding to the F_0 itself (e.g., Licklider, 1951; Schouten et al., 1962; Plack et al., 2005). The cochlea is the stage of the auditory system where the frequency-to-place mapping first occurs (tonotopic organization). Here, each harmonic component will mostly excite a specific region of the basilar membrane. Since the auditory-filter bandwidth is known to increase with increasing center frequency (Glasberg and Moore, 1990), while the spacing between neighbouring harmonics is constant and equal to the F_0 , each low-numbered harmonic will be processed within a distinct auditory filter, thus producing peaks and dips in the excitation pattern of the basilar membrane, while neighboring high-numbered harmonics will interact within the same auditory filter giving rise to a smooth excitation pattern (Fig. 1.1; Plack, 2005). As a result, the frequency of the individual harmonics can be retrieved from the sinusoidal pattern of vibration elicited by the low-numbered harmonics at specific places along the basilar membrane, while no place information is conveyed by the complex pattern of vibration elicited by the high-numbered harmonics. Because of the different separation along the basilar membrane, the low-numbered harmonics are said to be *resolved*, and the high-numbered harmonics are referred to as *unresolved*. Throughout this thesis, complex tones containing either resolved or unresolved harmonics will be referred to as resolved or unresolved complex tones, respectively.

The resolvability of a complex tone seems to depend on the lowest harmonic number present in the stimulus rather than on the frequency of the harmonic components per se (Bernstein and Oxenham, 2003; Plack et al., 2005). For example, a complex tone filtered in a high-frequency region (e.g., between 1.5 and 3.5 kHz)

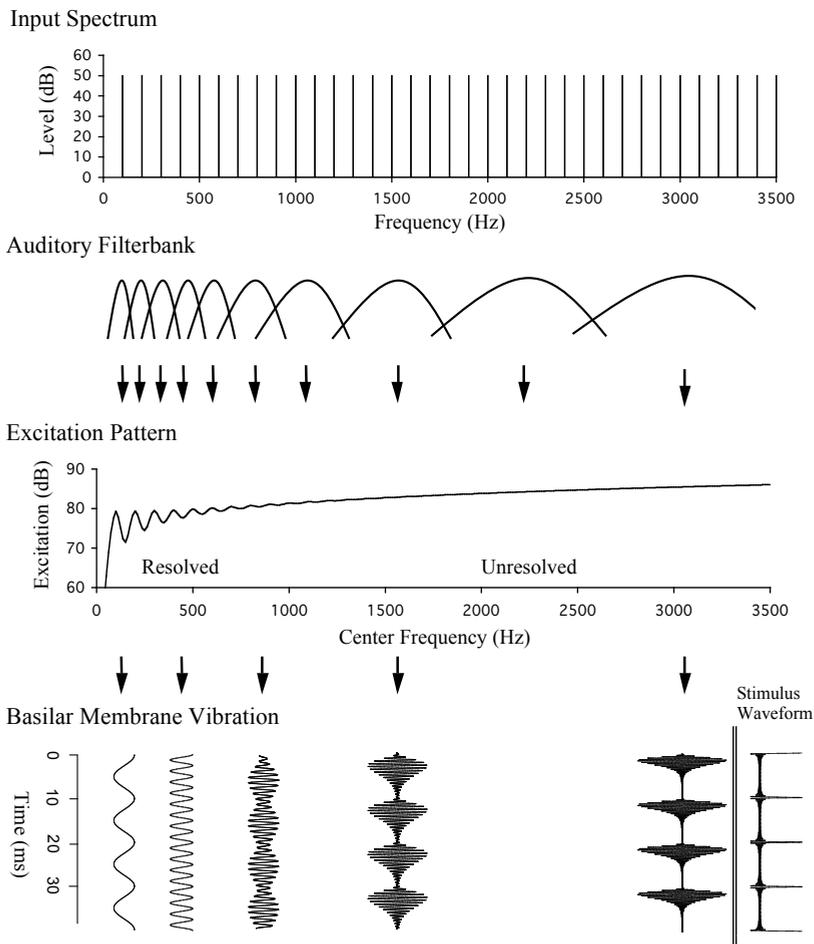


Figure 1.1: Schematic of the basilar membrane excitation pattern and pattern of vibration in response to a complex tone. Reproduced from Plack et al. (2014), with permission.

will be completely resolved by the basilar membrane at fairly large F_0 s (e.g., above about an F_0 of 300 Hz, Bernstein and Oxenham, 2006b), i.e., when the lowest harmonic number is below the 6th. In contrast, it will contain only unresolved harmonics at small F_0 s (e.g., $F_0 = 125$ Hz), i.e., when the lowest harmonic number is above about 12. Although there is no clear consensus on the harmonic number at which the transition from resolved to unresolved harmonics occurs, experimental results suggest that the transition occurs somewhere between the 5th and the 10th harmonic (for a review, see Plack et al., 2005; Moore and Gockel, 2011).

1.3 Complex-tone pitch coding mechanisms

Pitch perception and its underlying coding mechanisms have been investigated for decades to understand what information is necessary for the human auditory system to extract pitch (for a review, see De Cheveigné, 2005). Although some classical studies favored either a place-based (e.g., Ohm, 1843; Helmholtz, 1877; Goldstein, 1973; Wightman, 1973; Terhardt, 1974) or a temporal approach (e.g., Rutherford, 1886; Licklider, 1951; Licklider, 1959), more recent investigations suggest that both types of cues may be important for pitch coding (e.g., Shamma and Klein, 2000; Gockel et al., 2001; Heinz et al., 2001; Moore, 2003; Cedolin and Delgutte, 2005; Cedolin and Delgutte, 2010; Oxenham et al., 2011). Numerous studies have focused on the pitch coding mechanisms underlying pitch perception of resolved and unresolved complex tones (Hoekstra and Ritsma, 1977; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Grimault et al., 2002; Grimault et al., 2003; Moore and Moore, 2003; Bernstein and Oxenham, 2003; Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b; Moore et al., 2006a; Moore et al., 2006b; Bernstein and Oxenham, 2008; Moore and Glasberg, 2011). Different coding mechanisms have been suggested for complex tones containing either low-numbered resolved harmonics or high-numbered unresolved components. In fact, since resolved complex tones convey both the place and the time information of the individual harmonics along the basilar membrane, the pitch may be retrieved by comparing the phase-locking

pattern across characteristic frequency (obtained either via excitation pattern cues and/or temporal fine structure cues of the individual harmonics) with a set of harmonic templates (Shamma and Klein, 2000). Alternatively, the temporal pattern of neural spikes can be used to retrieve pitch (e.g., via autocorrelation; Licklider, 1951). In contrast, the pitch of unresolved complex tones can only be retrieved by the temporal information conveyed by envelope coding and/or the waveform fine structure information (although temporal fine structure seems not to be involved for harmonics above the 14th, Moore and Moore, 2003).

Some indirect evidence for different pitch-coding mechanisms for resolved and unresolved complex tones comes from studies on F_0 discrimination. First of all, resolved complex tones were found to elicit a more salient pitch percept than unresolved complex tones, whereby the F_0 -discrimination performance was more accurate when the complex tone contained resolved harmonics as compared to complexes with only unresolved harmonics (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; Bernstein and Oxenham, 2006a). However, Bernstein and Oxenham (2003) showed that increasing the resolvability of high-numbered harmonics by presenting the odd harmonics to one ear and the even harmonics to the other did not improve performance. Thus, the lowest harmonic number (or the harmonic rank), rather than resolvability per se, seemed to determine the transition between two different pitch-coding mechanisms, one based on the presence of spectro-temporal cues and one based on the temporal information carried by the high-numbered harmonics (Bernstein and Oxenham, 2003).

Additionally, studies on the effect of selective training on pitch-discrimination showed that learning was partially specific to the resolvability of the stimulus. In fact, listeners trained with a resolved complex tone showed larger improvements in F_0 -discrimination performance for another resolved complex tone than for an unresolved complex tone (Grimault et al., 2002; Carcagno and Plack, 2011). However, it is unclear whether these findings suggest the presence of separate mechanisms for pitch coding or simply support the hypothesis that the mechanisms for learning are partly stimulus-specific (specificity of learning, Carcagno and Plack, 2011).

1.4 Pitch perception and hearing loss

Sensorineural hearing loss (SNHL) is commonly associated with reduced frequency selectivity (Glasberg and Moore, 1986) and a reduced ability to extract temporal fine structure information (Moore et al., 2006b; Hopkins and Moore, 2007). Several studies reported that hearing-impaired (HI) listeners have disrupted abilities in pitch discrimination of complex tones (Hoekstra and Ritsma, 1977; Moore and Glasberg, 1988; Moore and Peters, 1992; Arehart, 1994; Arehart and Burns, 1999; Moore and Moore, 2003; Bernstein and Oxenham, 2006b). In particular, a poor pitch-discrimination performance was typically observed in HI listeners when the complex tones contained low-numbered harmonics (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Arehart, 1994; Moore and Glasberg, 1990; Moore and Peters, 1992; Bernstein and Oxenham, 2006b). For example, Moore and Peters (1992) found that the performance of HI listeners was worse when a complex tone with a low F_0 (below 200 Hz) contained low-numbered harmonics (1 to 12) than for complex tones without low harmonics (6 to 12). Thus, while the dominant harmonics for pitch perception are assumed to be between the first and the fifth in NH listeners (Plack et al., 2005), adding low-numbered harmonics led to a worse performance in HI listeners. Since the broadening of auditory filters in HI listeners leads to an increased number of unresolved harmonics as compared to NH listeners, it seems plausible that HI listeners rely more on the temporal information conveyed by the unresolved harmonics, rather than on the fine spectro-temporal information conveyed by the resolved harmonics (Moore and Carlyon, 2005). Supporting this hypothesis, it has been found that the performance of HI listeners is not always reduced as compared to NH listeners (Moore et al., 1998). In fact, some studies showed a similar performance of HI vs. NH listeners for pitch discrimination of unresolved complex tones (Arehart, 1994; Bernstein and Oxenham, 2006b). Thus, while fine spectro-temporal cues are disrupted, temporal envelope cues may be relatively more robust due to the presence of more harmonics interacting within the same filter and giving rise to a peakier envelope at the output of the filter. As a consequence of this, the relative importance of spectral and temporal cues for pitch processing may be altered in listeners with SNHL.

1.5 Pitch perception and musical experience

Since most sounds produced by musical instruments are harmonic complex tones, musicians are more trained than non-musicians to listen for and retrieve the pitch of such complex sounds. Possibly as a result of training, musicians have been shown to be more sensitive than non-musicians to discriminate fine spectral changes between complex tones. In fact, trained musicians were found to have two to six times better performance than non-musicians in complex-tone pitch discrimination (Spiegel and Watson, 1984; Micheyl et al., 2006; Allen and Oxenham, 2014; Bianchi et al., 2016b). This benefit was shown to depend on the overall duration of musical training, on the age when musical education started, as well as on the family of played instruments, with a smaller benefit for the musicians playing keyboard instruments as compared to strings and winds (Spiegel and Watson, 1984; Micheyl et al., 2006).

There seems to be a growing consensus on the role of a training-dependent plasticity for the enhanced performance of musicians (for a review see, Pantev and Herholz, 2011; Zatorre and Zarate, 2012). Many studies reported that musical training led to both anatomical and functional changes in the musicians' cortical and subcortical structures (e.g., Pantev et al., 1998; Zatorre, 1988; Schneider et al., 2002; Pantev et al., 2003; Musacchia et al., 2007; Wong et al., 2007; Foster and Zatorre, 2010; Herholz and Zatorre, 2012). Structural changes in the musicians' brain have been found in terms of an increase of gray-matter concentration and cortical thickness in motor and auditory-related areas (Schneider et al., 2002; Bermudez and Zatorre, 2005), as well as in frontal regions (Sluming et al., 2002) and on the volume of corpus callosum (Schlaug et al., 2009). Some studies (Hyde et al., 2009; Schlaug et al., 2009; Foster and Zatorre, 2010) showed that the degree of these structural changes was correlated with performance during musically-relevant behavioral tasks. Additionally, these structural changes were specific to the right auditory cortex (Hyde et al., 2009; Foster and Zatorre, 2010), consistent with the right auditory cortex being specialized in fine pitch processing (e.g., Zatorre, 1988; Johnsrude et al., 2000; Zatorre and Belin, 2001; Zatorre et al., 2002; Hyde et al., 2008).

Among the first investigations concerning functional changes in musicians, Pantev et al. (1998) reported an enhanced cortical activation to piano notes in musicians relative to a control group of non-musicians. This enhancement was not only correlated with the age at which musicians started musical training but was also seen to be specific to the timbre of the played instrument (Pantev et al., 2001). Additionally, electrophysiological studies observed an increased neural synchrony to music and speech in musicians already at a subcortical level, resulting in a more precise temporal and spectral representation of the signal (Musacchia et al., 2007; Wong et al., 2007; Parbery-Clark et al., 2009; Anderson and Kraus, 2011). Finally, two recent studies suggested that the musicians' advantage may be of peripheral origin. Psychoacoustical and physiological estimates of peripheral frequency selectivity suggested that musicians may have narrower auditory filters as compared to non-musicians (Bidelman et al., 2014; Bidelman et al., 2016). However, other studies showed no difference in peripheral frequency selectivity between musicians and non-musicians (Fine and Moore, 1993; Oxenham et al., 2003).

Although there is substantial evidence of changes in the musicians' neuroanatomy that are predictive of the musicians' enhanced performance, musical training alone cannot always account for these structural and functional changes (Foster and Zatorre, 2010). It cannot be excluded that anatomical predispositions might also play a role for the musicians' enhanced abilities to retrieve pitch (Zatorre and Zarate, 2012).

1.6 Overview of the thesis

This work focuses on complex-tone pitch perception and its objective representations at different stages along the auditory system in the normal and hearing-impaired population, as well as on how musical experience leads to changes in the perception of pitch. The research projects presented throughout the chapters of this thesis tackle the fundamental question about the nature of pitch-relevant cues in the human auditory system, by comparing pitch perception across different cohorts of listeners, specifically those showing enhanced pitch cues (e.g., musicians)

and those typically having disrupted pitch cues (i.e., hearing-impaired listeners) relative to a normal-hearing cohort of non-musically trained listeners.

In *Chapter 2*, pitch-discrimination performance is behaviorally estimated in normal-hearing and hearing-impaired listeners for complex tones with harmonics added either in sine or random phase. When the harmonics are unresolved, the difference in pitch-discrimination performance between the random-phase and the sine-phase condition can be considered as an estimate of temporal envelope processing. This difference in performance is compared between the two groups of listeners to assess whether changes in the temporal envelope representation occur in the hearing-impaired listeners. Cochlear compression and frequency selectivity are additionally estimated in the same listeners to clarify whether changes in envelope cues of hearing-impaired listeners can be ascribed to their cochlear damage. Cochlear compression and frequency selectivity are finally considered in a simplified peripheral model to clarify their relative contribution to a possible envelope enhancement following SNHL.

In *Chapter 3*, the use of task-evoked pupil dilation is investigated as an objective measure of effort during a pitch-discrimination task in normal-hearing listeners. Pupil responses are recorded while listeners perform a pitch-discrimination task with complex tones of varying pitch salience. Since resolved complex tones are known to elicit a stronger pitch percept than unresolved complex tones, it is hypothesized that task-evoked effort would increase with decreasing the salience of the evoked pitch percept.

Chapter 4 investigates the perceptual enhancement of musicians in pitch-discrimination for resolved and unresolved complex tones. A first behavioral experiment is performed to clarify whether the musicians' advantage for complex-tone pitch discrimination occurs for both resolved and unresolved complex tones. A second experiment uses pupil responses to compare the effort of musicians and non-musicians while performing a pitch-discrimination task with complex tones of varying resolvability. The aim of this study is to clarify whether the enhanced performance in musicians can be ascribed to an increased peripheral frequency selectivity and/or to a different processing effort in performing the task.

In *Chapter 5*, the perceptual enhancement of musicians in pitch-discrimina-

tion is further investigated via functional magnetic resonance imaging (fMRI). A first behavioral experiment is performed to estimate the individual pitch-discrimination thresholds for both resolved and unresolved complex tones. In a second experiment, cortical neural responses are measured via fMRI in musicians and non-musicians, while the participants are asked to perform a pitch-discrimination task that is adjusted in difficulty according to their behavioral thresholds. The aim of this study is to clarify whether musicians show an increased cortical activation in response to resolved and unresolved complex tones as compared to non-musicians, despite the difficulty of the task being adjusted across participants. It is hypothesized that an increased activation in musicians may indicate the existence of functional changes, possibly as a result of musical training, which may be specific to either the resolved or unresolved harmonics.

Finally, the main findings of each chapter are summarized and discussed in *Chapter 6*. Implications for temporal envelope coding in hearing-impaired listeners as well as the effects of musical training on pitch discrimination of resolved and unresolved complex tones are further discussed.

In the *Appendix*, a different topic of hearing research is addressed. The perceptual phenomenon of dominance of the directional information contained in the first arriving sound, referred to as the precedence effect, is investigated behaviorally and objectively. The contribution of peripheral versus central auditory processes to the precedence effect is investigated by comparing physiological (otoacoustic emissions and auditory brainstem responses) and psychoacoustical data in normal-hearing listeners.

2

Complex-tone pitch discrimination in listeners with sensorineural hearing loss^a

Abstract

Physiological studies have shown that noise-induced sensorineural hearing loss (SNHL) enhances the amplitude of envelope coding in auditory-nerve fibers. As pitch coding of unresolved complex tones is assumed to rely on temporal envelope coding mechanisms, this study investigated pitch-discrimination performance in listeners with SNHL. Pitch-discrimination thresholds were obtained for 14 normal-hearing (NH) and 10 hearing-impaired (HI) listeners for sine-phase (SP) and random-phase (RP) complex tones. When all harmonics were unresolved, the HI listeners performed, on average, worse than NH listeners in the RP condition but similarly to NH listeners in the SP condition. The increase in pitch-discrimination performance for the SP relative to the RP condition (F_0 DL ratio) was significantly larger in the HI as compared to the NH listeners. Cochlear compression and auditory-filter bandwidths were estimated in the same listeners. The estimated reduction of cochlear compression was significantly correlated with the increase in the F_0 DL ratio, while no correlation was found with filter bandwidth. The effects of degraded frequency selectivity and loss

^a This chapter is based on Bianchi, F., Fereczkowski, M., Zaar, J., Santurette, S., Dau, T. (in press), *Trends in Hearing*.

of compression were considered in a simplified peripheral model as potential factors in envelope enhancement. The model revealed that reducing cochlear compression significantly enhanced the envelope of an unresolved SP complex tone, while not affecting the envelope of a RP complex tone. This envelope enhancement in the SP condition was significantly correlated with the increased pitch-discrimination performance for the SP relative to the RP condition in the HI listeners.

2.1 Introduction

Pitch perception and its underlying coding mechanisms have been investigated for decades to understand what information is necessary for the human auditory system to extract pitch (for a review, see De Cheveigné, 2005). Although some studies favored either a place-based (e.g., Ohm, 1843; Goldstein, 1973; Wightman, 1973; Terhardt, 1974; Helmholtz, 1877) or a temporal approach (e.g., Licklider, 1951; Licklider, 1959; Rutherford, 1886), more recent investigations suggest that both types of cues may be important for pitch coding (e.g., Shamma and Klein, 2000; Gockel et al., 2001; Heinz et al., 2001; Moore, 2003; Cedolin and Delgutte, 2005; Cedolin and Delgutte, 2010; Oxenham et al., 2011).

Numerous studies have focused on the pitch coding mechanisms underlying pitch perception of complex tones (Hoekstra and Ritsma, 1977; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Moore and Moore, 2003; Bernstein and Oxenham, 2003; Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b; Moore et al., 2006a; Moore et al., 2006b; Bernstein and Oxenham, 2008; Moore and Glasberg, 2011). Different coding mechanisms were suggested for complex tones containing either low-numbered resolved harmonics or high-numbered unresolved components. While resolved components are processed by separate auditory filters and produce distinct ripples in the excitation pattern, neighboring unresolved components are processed within the same auditory filter and their interaction gives rise to a smooth excitation pattern which does not convey place information from

which the frequency of individual harmonics can be retrieved (Plack, 2005). As a result, the pitch of resolved complex tones may be retrieved by fine spectral and/or temporal cues, while the pitch of unresolved complex tones can only be retrieved by the temporal information conveyed by envelope coding.

Sensorineural hearing loss (SNHL) is commonly associated with reduced frequency selectivity (Glasberg and Moore, 1986) and a reduced ability to extract temporal fine structure information (Moore et al., 2006b; Hopkins and Moore, 2007). However, recent physiological studies in animals showed that noise-induced SNHL increases the temporal precision and the amplitude of envelope coding in single auditory-nerve fibers (Kale and Heinz, 2010; Henry et al., 2014). These findings were ascribed to a variety of factors, such as broader auditory filters, a reduction of cochlear compression due to outer hair cell damage and altered auditory-nerve response temporal dynamics (Scheidt et al., 2010). Thus, while fine spectro-temporal cues are disrupted, temporal envelope cues may be enhanced and the relative importance of spectral and temporal cues for pitch processing may be altered in listeners with SNHL. Although several studies reported that hearing-impaired (HI) listeners have disrupted abilities in pitch discrimination of complex tones (Hoekstra and Ritsma, 1977; Moore and Glasberg, 1988; Moore and Peters, 1992; Arehart, 1994; Arehart and Burns, 1999; Moore and Moore, 2003; Bernstein and Oxenham, 2006b), it has been found that the performance of HI listeners is not always disrupted as compared to NH listeners (Moore et al., 1998).

In fact, while most studies reported a degraded performance of HI listeners in pitch discrimination of stimuli containing low-order harmonics (Hoekstra and Ritsma, 1977; Hoekstra, 1979; Arehart, 1994; Moore and Glasberg, 1990; Moore and Peters, 1992; Bernstein and Oxenham, 2006b), which may be related to a reduced frequency selectivity (Bernstein and Oxenham, 2006b; Moore and Glasberg, 2011), some studies showed a similar performance of HI vs. NH listeners for pitch discrimination of unresolved complex tones and also a comparable performance of HI listeners for pitch discrimination of resolved vs. unresolved stimuli (Arehart, 1994; Bernstein and Oxenham, 2006b). Since the broadening of auditory filters in HI listeners leads to an increased number of unresolved harmonics as compared to NH listeners, it seems plausible that HI listeners rely more on the temporal information

conveyed by the unresolved harmonics, rather than on the fine spectro-temporal information conveyed by the resolved harmonics (Moore and Carlyon, 2005). It is still unclear whether the altered importance of temporal vs. spectral cues for pitch discrimination may be additionally due to the suggested enhancement of temporal envelope coding with SNHL (Kale and Heinz, 2010; Henry et al., 2014).

The aim of the present behavioral study was to clarify: i) whether human listeners with SNHL show an enhancement of temporal envelope coding, ii) if this enhancement is related to the broadening of auditory filters and/or to the reduction of cochlear compression, and iii) how this enhancement affects pitch discrimination of complex tones. Pitch discrimination of complex tones was investigated behaviorally as a function of the fundamental frequency (F_0) in NH listeners and listeners with SNHL (Experiment I). Additionally, an amplitude-modulation detection experiment (Experiment II) was performed in the same listeners to assess temporal envelope coding abilities and to estimate individual auditory-filter bandwidths based on detectability of the modulation sidebands. Furthermore, the basilar-membrane input/output function (BM I/O) was estimated for the HI listeners using a forward-masking task (Experiment III), to assess the role of degraded cochlear compression for pitch discrimination of unresolved complex tones. Finally, a simplified peripheral model, adjusted according to the auditory-filter bandwidth and cochlear-compression estimates obtained in Experiments II and III, was used to clarify the role of degraded cochlear compression and filter broadening for pitch-discrimination performance based on the envelope peakiness of the unresolved complexes at the output of the filter.

While in previous studies (e.g., Hoekstra, 1979; Glasberg and Moore, 1989; Moore and Glasberg, 1990; Moore and Peters, 1992; Bernstein and Oxenham, 2006b; Moore and Glasberg, 2011) the individual performance in pitch discrimination was correlated with individual measures of frequency selectivity, the novelty of the current study is that pitch discrimination was further investigated as a potential indicator of temporal envelope processing, on which pitch coding of unresolved complex tones is assumed to rely.

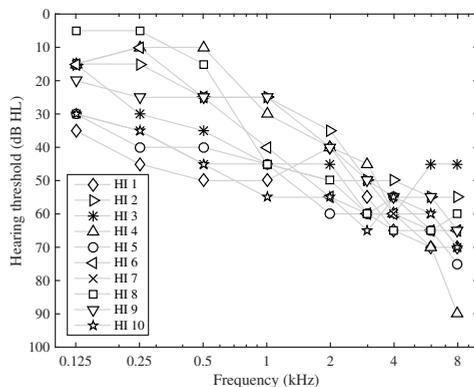


Figure 2.1: Hearing thresholds in the test ear for the 10 HI listeners who participated in this study. The thresholds were obtained via conventional audiometry.

2.2 Methods

2.2.1 Listeners and experimental setup

Fourteen NH listeners (6 females), aged from 22 to 28 years old, and ten HI listeners (4 females), aged from 65 to 81 years old, participated in this study. All NH listeners had hearing thresholds of less than 20 dB hearing level (HL) at all audiometric frequencies between 125 Hz and 8 kHz. The HI listeners had hearing thresholds between 25 and 65 dB HL at the audiometric frequencies between 1 and 4 kHz. The individual hearing thresholds of the HI listeners are reported in Fig. 2.1 and the hearing thresholds at 2 kHz are listed in Table 2.1. All experiments were carried out monaurally, whereby the NH listeners were tested at their right ear and the HI listeners at their best ear matching the inclusion criteria. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark.

2.2.2 Experiment I: Pitch-discrimination of complex tones

The ability to discriminate the pitch of resolved and unresolved complex tones was assessed via difference limens for fundamental frequency (F_0 DLs) as a function of

F_0 .

Procedure

A three-alternative forced choice (3-AFC) paradigm was used in combination with a weighted up-down method (Kaernbach, 1991) to measure the 75% point on the psychometric function. For each trial, two intervals contained a reference complex tone with a fixed fundamental frequency ($F_{0,\text{ref}}$: 100, 125, 150, 175, 200, 250, 300, 400, 500 Hz) and one interval contained a deviant complex tone with a larger F_0 ($F_{0,\text{dev}}$). The initial difference in F_0 between reference and deviant, ΔF_0 , $(F_{0,\text{dev}} - F_{0,\text{ref}}) / F_{0,\text{ref}}$, was set to 20% and was then decreased by a varying step size every second reversal. After each correct answer, ΔF_0 was decreased by a factor of 2.23 until the first reversal, by a factor of 1.7 until the third reversal and by a factor of 1.16 for the following seven reversals. For each run, $F_{0,\text{ref}}$ was roved from trial to trial from a $\pm 5\%$ uniform distribution around the nominal value. A random level perturbation of ± 2.5 dB was applied to each interval, to reduce potential loudness cues. The listener's task was to select the interval containing the tone with the highest pitch. The threshold for each run was obtained as the geometric mean of the last six reversals. Before the actual test, the listeners performed three repetitions as training. The final value of $F_0\text{DL}$ was calculated from the mean of three repetitions.

Stimuli

All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 300-ms complex tones embedded in threshold equalizing noise (TEN, Moore et al., 2000). For the NH listeners, the sound pressure level (SPL) of the TEN was set to 55 dB per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) to mask the combination tones. For the HI listeners, pure tone detection in quiet was performed at 1.5, 2 and 3 kHz (2 repetitions per frequency) and the level of the TEN was set at the maximum threshold measured in this range. The complex tones were created by summing harmonic components either in sine phase (SP) or random phase (RP) to vary the envelope peakiness. Summing the harmonics in

SP yields to a peaky signal envelope, while summing the harmonics in RP yields to a much flatter envelope. All HI listeners were tested in the SP and RP conditions, whereas only nine out of the 14 NH listeners completed the measurements for both conditions. Conditions of varying resolvability were achieved by bandpass filtering the complexes in a high-frequency region (HF, 1500-3500 Hz), with 50 dB/octave slopes, and by varying the F_0 (Bernstein and Oxenham, 2006b). In order to keep the sensation level (SL) of the complex tones approximately constant across listeners, pure tone detection in TEN background was performed at 1.5, 2, and 3 kHz (three repetitions per frequency). For each listener, the mean detection threshold was calculated across the three frequencies and the level of each component of the complex tone was set at 12.5 dB SL re the mean threshold (obtained levels for each listener are presented in Table 2.1). The sound stimuli were delivered through headphones (Sennheiser HDA 200).

2.2.3 Experiment II: Amplitude-modulation detection

The temporal modulation transfer function (TMTF), i.e., the amplitude-modulation detection threshold as a function of the modulation frequency (f_m), was estimated for a 2-kHz carrier. This measure yielded estimates of two quantities: amplitude-modulation detection and auditory-filter bandwidth at 2 kHz. For each listener, the auditory-filter bandwidth was estimated as the f_m for which the sidebands became resolved. Five out of the 14 NH and all 10 HI listeners participated in this experiment.

Procedure

A 3-AFC paradigm, in combination with a weighted up-down rule, was used to measure modulation detection thresholds at the 75% point of the psychometric function. For each trial, two intervals contained a pure tone at 2 kHz and one interval contained a sinusoidally amplitude-modulated 2-kHz sinusoid modulated at $f_m = 25, 50, 100, 150, 200, 300, 400, 500, 800, 1000, \text{ or } 1500$ Hz. The initial modulation depth (20 log m) was set to -10 dB and was then adaptively varied in dB steps with starting and ending values of 5 and 1 dB, respectively. For each

interval, the carrier frequency was roved from a $\pm 3\%$ uniform distribution around 2 kHz. A random level perturbation of ± 1.5 dB was applied to each interval to minimize loudness cues. The listener's task was to select the interval containing the modulated tone. The threshold for each run was obtained as the geometric mean of the last six reversals. Before the actual test, the listeners performed one repetition as training. The final threshold was calculated from the mean of three repetitions. For each listener, the auditory filter bandwidth was estimated as the fm leading to a modulation threshold that was 10.5 dB below the maximum point of the TMTE. This point was selected since it led to an estimated filter bandwidth of 325 Hz at 2 kHz for NH listeners, which corresponds to the mean equivalent rectangular bandwidth (ERB) estimated via the notched-noise method by Bernstein and Oxenham (2006b).

Stimuli

All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 300-ms pure tones. The carrier level was set to the same level as the nominal components of the complex tones in the pitch discrimination experiment (i.e., at 12.5 dB SL *re* the TEN level used in experiment I, see Table 2.1). No background noise was used. The stimuli were presented via Sennheiser HDA 200 headphones.

2.2.4 Experiment III: Estimates of BM I/O function and cochlear compression

The residual peripheral compression was estimated in nine out of the 10 HI listeners (all except HI 7) by estimating the individual BM I/O functions at 2 kHz. The BM I/O functions were derived from the temporal masking curves (TMCs) measured via a forward masking experiment for the nine listeners.

Procedure

Masker thresholds were measured as a function of the temporal gap between a 2-kHz probe and a masker tone, either “on-frequency” at 2 kHz or “off-frequency” at 0.6 times the probe frequency. The thresholds were tracked using the Grid method (Fereczkowski, 2015), which reduces the duration of the forward-masking experiment. After three repetitions of the measurement, the on-frequency thresholds were fitted for each listener with either two- or single sections, depending on the estimated value of the Bayesian Information Criterion (Schwarz, 1978). This criterion was used to avoid model overfitting. Off-frequency thresholds were fitted with single sections in all cases. The fits were used to infer BM I/O functions following the paradigm of Nelson et al. (2001). The inverse slope of the section comprising the input stimulus level was taken as an estimate of the compression ratio (CR) at 2 kHz.

Stimuli

The masker tone duration was 200 ms and the probe tone duration was 16 ms. Both were gated with 4-ms raised-cosine onset and offset ramps, hence the lengths of the steady state portions were 192 and 8 ms, respectively. The probe level was set at 10 dB above the absolute probe threshold. The stimuli were generated in MATLAB (44100 Hz sampling rate, 24-bit rate) and presented via Sennheiser HDA 200 headphones.

2.2.5 Modeling the effects of cochlear compression and frequency selectivity on envelope peakiness

HF-filtered complex tones ($F_0 = 100$ Hz) with harmonics added either in SP or RP were passed through a single fourth-order gammatone filter centred at 2 kHz, which was adjusted in bandwidth to an “average NH listener” as well as to the individual HI listeners according to the estimates from Experiment II (listed in Table 2.1). The signal at the output of the filter (S_{flt}) was then compressed according to $S_{\text{comp}} = \text{sign}(S_{\text{flt}}) \cdot |S_{\text{flt}}|^{\frac{1}{\text{CR}}}$, where CR denotes the individual compression ratios estimated

Table 2.1: Individual values of hearing level at 2 kHz, sound pressure level per harmonic component used in Experiment I, auditory filter bandwidth estimated from Experiment II and compression ratio (CR) estimated from Experiment III for the mean of the NH listeners and the individual HI listeners. (*) from Lopez-Poveda et al. (2003).

Listener	Hearing level at 2 kHz (dB HL)	Component level (dB SPL)	Auditory filter BW (Hz)	CR
Mean NH	< 20	65	325	6.0 (*)
HI 1	40	71.2	898	1.3
HI 2	35	68.5	646	2
HI 3	45	73.8	753	1.7
HI 4	40	71	587	2.9
HI 5	60	80	979	1.4
HI 6	55	73	915	0.8
HI 7	50	72	1390	N/A
HI 8	50	77.2	968	1.4
HI 9	40	70	577	2.3
HI 10	55	80	778	1

from Experiment III (see Table 2.1). The Hilbert envelope of the compressed signal was obtained and band-limited using a first-order low-pass filter with a cut-off frequency of 150 Hz (Kohlrausch et al., 2000; Ewert and Dau, 2000). As a descriptor of the peakiness of the resulting envelope E , the modulation power P_{mod} of the output signal was calculated as the ratio between the envelope power and the envelope DC, $P_{mod} = \frac{\frac{1}{N} \sum_{n=1}^N E(n)^2}{\left(\frac{1}{N} \sum_{n=1}^N E(n)\right)^2}$, where N denotes the number of samples. The simulations for the RP condition were iterated 100 times in order for the random process to converge. The obtained P_{mod} values were then averaged across iterations. As a result, P_{mod} values were obtained for a NH profile and the individual HI profiles except for HI 7 (cf. Table 2.1). For each auditory profile, the modulation power was obtained for the SP ($P_{mod, SP}$) and RP ($P_{mod, RP}$) complex tones. Finally, the modulation power ratio, $\frac{P_{mod, SP}}{P_{mod, RP}}$, was calculated.

2.3 Results

2.3.1 Experiment I: Pitch-discrimination of complex tones

Figure 2.2 (top panels) depicts the mean pitch-discrimination thresholds for NH listeners (black solid symbols), as well as the individual thresholds for HI listeners (open symbols), for the SP condition (left panel), the RP condition (middle panel) and the ratio between the RP and the SP thresholds (right panel). The thresholds for the SP and RP conditions showed similar trends for the NH listeners, whereby F_0 DLs decreased with increasing F_0 . A mixed-model ANOVA on the log-transformed F_0 DLs with F_0 and phase as fixed effects and subjects as a random effect confirmed a significant effect of F_0 for the NH listeners ($F(8,176) = 55.61$, $p < 0.001$), as well as a significant interaction of F_0 and phase ($F(8,176) = 3.05$, $p = 0.003$). These findings are in agreement with previously reported pitch-discrimination thresholds (e.g., Bernstein and Oxenham, 2006b), where the improvement in performance with increasing F_0 was thought to reflect the progressive increase of the resolvability of the harmonics and/or the increase in the effectiveness of temporal fine-structure cues (Moore et al., 2006). Additionally, the SP condition yielded lower thresholds as compared to the RP condition at low F_0 s (i.e., in the presence of unresolved harmonics). This benefit in performance for the SP condition relative to the RP condition (referred to as F_0 DL ratio, right panels in Fig. 2.2) was, on average, of about a factor of 1.4 for F_0 s below 200 Hz for the nine NH listeners who completed both measurements. No phase effects were found for F_0 s equal or larger than 200 Hz (mean F_0 DL ratio of 0.95), consistent with the presence of resolved harmonics in the NH listeners above this F_0 for complex tones filtered between 1.5 and 3.5 kHz (e.g., Bernstein and Oxenham, 2006b; Bianchi et al., 2016b).

The mean performance of the 10 HI listeners was generally worse than that of the NH listeners. In fact, although some HI listeners showed a better performance than the NH listeners at low F_0 s, the thresholds for the HI listeners were, on average, larger than the thresholds for NH listeners (see Fig. 2.2, left and middle bottom panels). A mixed-model ANOVA with F_0 , group and phase as fixed factors and listeners as a random factor nested in group confirmed a significant effect of the

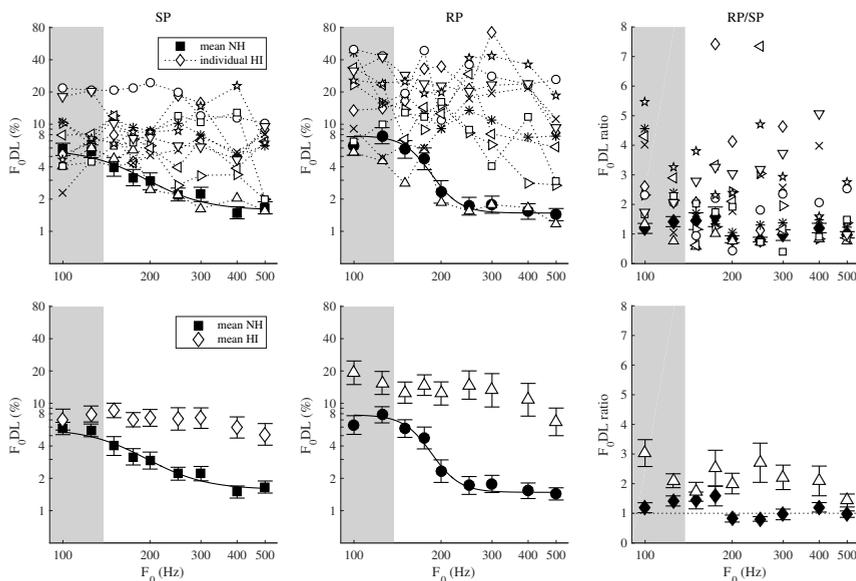


Figure 2.2: Pitch-discrimination thresholds for the SP condition (left panels) and RP condition (middle panels). The right panels depict the ratio of the RP and SP thresholds (F_0DL ratio). The solid symbols depict the mean results for 14 NH listeners in the left panels and 9 NH listeners in the middle and right panels. The open symbols depict the individual results (top panels, same symbols as in Fig. 2.1) and the mean results (bottom panels) for the 10 HI listeners. Error bars depict the standard error of the mean. The grey-shaded region highlights the conditions (at F_0 s of 100 and 125 Hz) for which the harmonics are considered to be unresolved.

fixed factors (F_0 : $F(8,335) = 29.18$, $p < 0.001$; group: $F(1, 335) = 22.25$, $p < 0.001$; phase: $F(1, 335) = 42.11$, $p < 0.001$), as well as a significant interaction of group and phase ($F(1, 335) = 39.22$, $p < 0.001$) and of group and F_0 ($F(8, 335) = 10.46$, $p < 0.001$). The grey-shaded area in Fig. 2.2 depicts the two conditions (at F_0 s of 100 and 125 Hz) for which the harmonics could be considered completely unresolved, i.e., when the lowest harmonic number was larger than or equal to 12 (Moore and Moore, 2003). For these two unresolved conditions a mixed-model ANOVA (fixed factors: group and F_0 ; listeners as a random factor nested in group) revealed no significant difference between the thresholds of the NH vs. the HI listeners for the SP condition (group effect: $F(1,47) = 1.53$, $p = 0.23$), while a significant difference

was present for the RP condition (group effect: $F(1,37) = 9.44$, $p = 0.007$). Two post-hoc one-tailed t-tests using Bonferroni-adjusted alpha levels of 0.025 revealed significantly larger thresholds for the HI vs. the NH listeners for both unresolved RP conditions (100 Hz: $p = 0.002$; 125 Hz: $p = 0.02$). Thus, these findings revealed that HI listeners performed similarly to NH listeners in pitch discrimination of unresolved complex tones for the SP condition and worse than NH listeners for the RP condition. Additionally, while NH listeners showed a moderate benefit in performance for the SP condition relative to the RP condition (mean F_0 DL ratio of 1.3 for the two unresolved conditions; right panels in Fig. 2.2), HI listeners showed a larger benefit, on average, of about a factor of 2.6 for the unresolved conditions.

2.3.2 Experiment II: Amplitude-modulation detection

Figure 2.3a depicts the amplitude-modulation detection thresholds for the individual HI listeners (open symbols), as well as the mean modulation thresholds for the five NH listeners who completed Experiment II (filled squares). The modulation thresholds for the NH listeners were independent of f_m up to a modulation rate of 200 Hz. At modulation rates above 200 Hz, the thresholds decreased with increasing f_m , due to detection of the resolved sidebands (Kohlrausch et al., 2000; Ewert and Dau, 2000). For the HI listeners, the TMTFs were flat up to modulation rates of about 100 Hz. At these low f_m s, thresholds for most of the HI listeners were lower than for the NH listeners, indicating a higher sensitivity to detect amplitude modulations. A one-way unbalanced ANOVA on the thresholds up to 100 Hz confirmed a significant group effect ($F(1,44) = 5.98$; $p = 0.019$, see mean thresholds on Figure 2.3b). Above 100 Hz, thresholds increased up to modulation rates of about 400 Hz (or higher for some HI listeners) due to central limitations of the auditory system to detect fast envelope fluctuations (Kohlrausch et al., 2000; Ewert and Dau, 2000). After the maximum point of the TMTE, the thresholds of the HI listeners decreased at different rates as the sidebands became resolved. The dotted vertical lines in Fig. 2.3a depict the individual filter bandwidths, estimated as the f_m leading to a modulation threshold (on the fitted curve) that was 10.5 dB below the maximum point of the TMTE. For the HI listeners, the estimated filter bandwidths ranged

from 577 Hz (HI 9) to 1390 (HI 7). The individual values are presented in Table 2.1.

2.3.3 Experiment III: Estimates of BM I/O function and cochlear compression

Figure 2.4 depicts the TMC thresholds (on-frequency masker: open symbols; off-frequency masker: filled circles) measured in nine HI listeners, together with the corresponding fits. The measured masking thresholds increased with increasing masker-probe gap, consistent with the TMC data reported in the literature (e.g., Nelson et al., 2001). For most listeners the fitted sections to the on-frequency TMCs (solid lines) were steeper than the corresponding off-frequency fits (dashed lines), while for other listeners (HI 6, HI 10), the on- and off-frequency fits showed similar slopes. This is consistent with some residual peripheral compression affecting the on-frequency maskers in case of the former listeners, but not the latter.

Figure 2.5 depicts the BM I/O functions (solid lines) estimated for the same nine listeners from the TMC fits. The linear reference is indicated by the dashed lines. The portions of the BM I/O functions that are shallower than the linear reference indicate the presence of peripheral compression in a given listener. The BM I/O functions represent the off-frequency TMC threshold on the ordinate (i.e., the BM output level) vs. the on-frequency TMC threshold on the abscissa (i.e., the BM input level) for each given masker-probe gap. Thus, as the BM I/O functions were estimated only in the range where both on- and off-frequency TMCs were measured, the obtained BM input-level range differed among listeners (i.e., from 12 dB for HI 3 to 34 dB for HI 1 and HI 4). The individual peripheral compression at 2 kHz was estimated as the inverse of the slope (i.e., the compression ratio, CR, see Table 2.1) of the fitted section comprising the input stimulus level (depicted by the asterisks in Fig. 2.5). This level was estimated for each listener as the overall level of a HF-filtered complex tone (at $F_0 = 100$ Hz), at the output of an individually adjusted gammatone filter centered at 2 kHz.

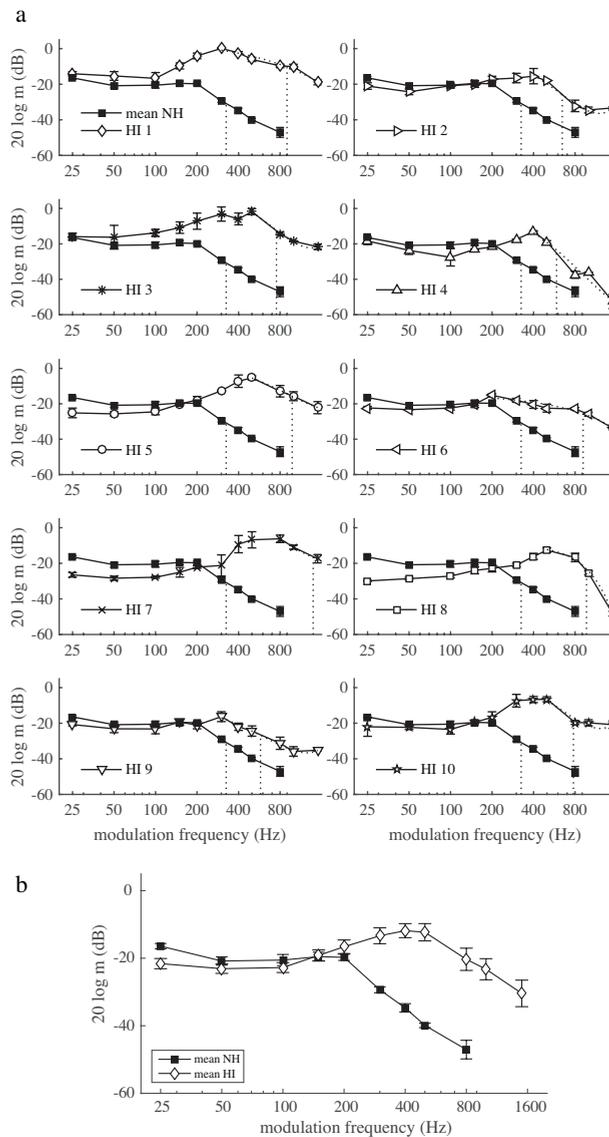


Figure 2.3: a) Amplitude-modulation detection thresholds for a 2-kHz sinusoidal carrier as a function of the modulation frequency for the 10 HI listeners (same open symbols as in Fig. 2.1; error bars depict the standard deviation across the three repetitions of each experimental condition). The mean thresholds for five NH listeners are also depicted in each panel for a comparison purpose (filled squares; error bars depict the standard error of the mean). The dashed vertical lines depict the estimated filter bandwidth as the fm leading to a modulation threshold that was 10.5 dB below the maximum point of the TMTF (the obtained bandwidths are listed in Table 2.1). b) Mean thresholds for NH (closed squares) and HI (open diamonds) listeners. Error bars depict the standard error of the mean.

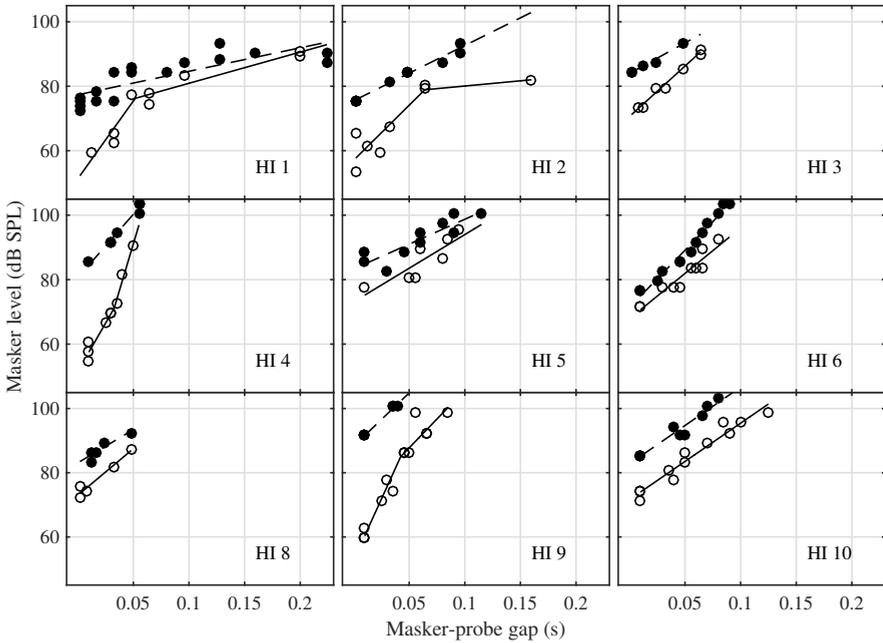


Figure 2.4: Temporal masking curves (TMCs) for nine HI listeners (HI 7 not measured), together with the corresponding fits. The on- and off-frequency thresholds are depicted with open and filled circles, respectively. The fits to the on-frequency data are shown with a solid line while the single-section fits to the off-frequency data are shown with a dashed line.

2.3.4 Effects of cochlear compression and frequency selectivity on pitch discrimination

As influencing factors such as musical training and individual cognitive resources, as well as individual limitations (e.g., neural synchrony, internal noise level) are likely to affect the overall pitch-discrimination performance, the ratio between the RP and SP thresholds (F_0DL ratio) was calculated for the individual HI listeners as well as for the mean of the NH listeners (Fig. 2.2, right panels). The F_0DL ratio quantifies the relative increase in pitch-discrimination performance for the unresolved SP complex tones with respect to their RP counterparts and allows for

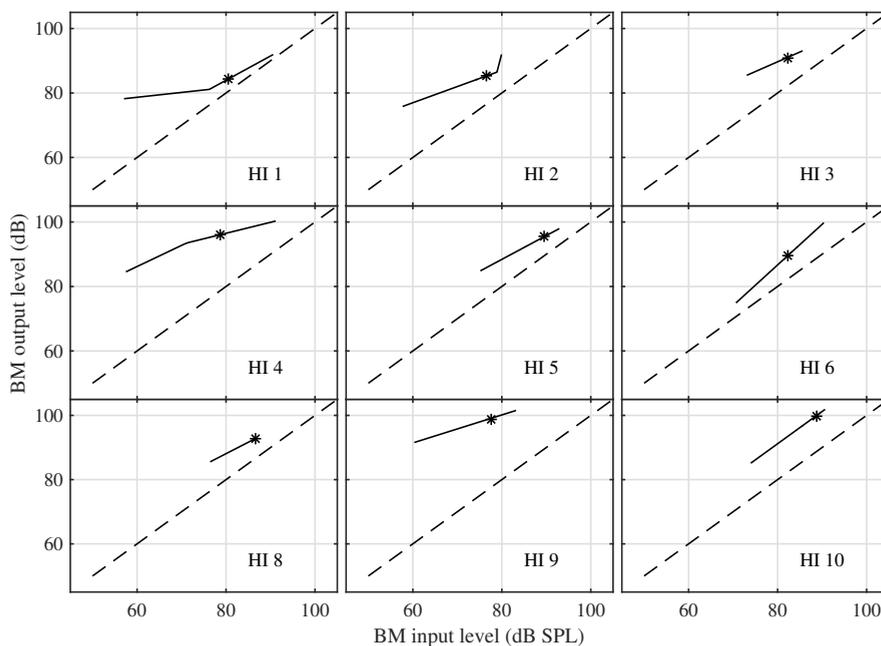


Figure 2.5: BM I/O functions (solid lines) estimated from the TMCs for nine HI listeners (HI 7 not measured). The dashed line depicts the linear reference, i.e., the BM I/O function assuming absent peripheral compression. The asterisks show the estimated levels of a HF-filtered complex tone at F_0 of 100 Hz at the output of individually adjusted auditory filters at 2 kHz. The peripheral compression was estimated at the levels marked by the asterisks and the individual values are listed in Table 2.1.

a comparison across listeners that is unbiased by the individual factors. Figure 2.6 shows the mean F_0 DL ratio for the two unresolved conditions (at F_0 s of 100 and 125 Hz) as a function of the estimated reduction of cochlear compression ($1/CR$, calculated from Experiment III at the level indicated by the asterisk in Fig. 2.5; left panel in Fig. 2.6) and filter bandwidth (estimates from Experiment II; right panel in Fig. 2.6). The increase of the F_0 DL ratios for the HI listeners was significantly positively correlated with the estimated loss of cochlear compression (left panel in Fig. 2.6: $R^2 = 0.56$, $p = 0.002$). Thus, the lower the residual cochlear compression the larger was the increase in pitch-discrimination performance for the SP relative

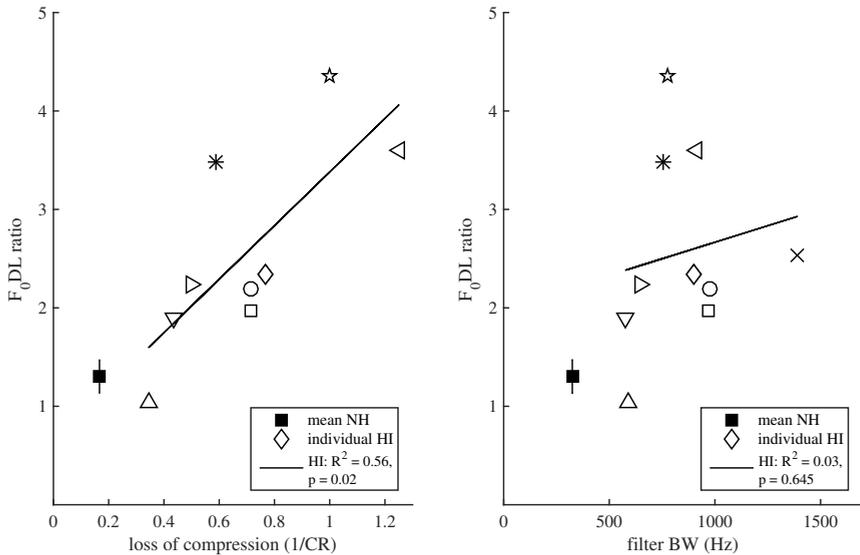


Figure 2.6: Mean F_0 DL ratios for the two unresolved conditions (at F_0 s of 100 and 125 Hz) as a function of the estimated loss of cochlear compression (left panel) and filter bandwidth (right panel). Solid symbols depict the mean results for the 9 NH listeners that measured both SP and RP conditions. The open symbols (same symbols as in Fig. 2.1) depict the individual results for HI listeners. Error bars depict the standard error of the mean. The correlations were carried out only across the data for the HI listeners.

to the RP complex tones. No significant correlation was found between F_0 DL ratio and auditory filter bandwidth ($R^2 = 0.03$, $p = 0.645$; right panel in Fig. 2.6). Overall, these findings suggest that loss of cochlear compression was the dominant factor increasing the pitch-discrimination performance for the unresolved SP complex tones relative to their RP counterparts.

2.3.5 Modeling the effects of cochlear compression and frequency selectivity on envelope peakiness

The left panels in Figure 2.7 depict the modulation power of the SP (open symbols) and RP (closed symbols) complex tones, estimated at the output of a peripheral model individually adjusted according to the auditory profiles of the nine HI and

the mean of the NH listeners. In the model, three simulations were run to clarify the relative effect of auditory-filter bandwidth and cochlear compression on the envelope representation of unresolved complex tones. In a first simulation (top panels), auditory-filter bandwidth was varied according to the estimates from Experiment II, while cochlear compression was fixed at a common value for NH listeners ($CR = 6$, Lopez-Poveda et al., 2003). The simulation revealed no effect of filter bandwidth on the modulation power of either the SP or RP signals. In a second simulation (middle panels), cochlear compression was varied according to the estimates from Experiment III, while filter bandwidth was fixed at the value of 325 Hz estimated for NH listeners (Experiment II). Reducing cochlear compression yielded an increase in the modulation power of the SP complex tone, indicating an increase of the envelope peakiness, while hardly affecting the modulation power of the RP complex tones. In fact, since compression is a non-linear operation, it mainly reduces the modulation depth of peaky signals. Thus, a reduction of compression yielded a much larger enhancement of the modulation depth for the SP than for the RP stimuli. In a third simulation (bottom panels), both filter bandwidth and cochlear compression were varied according to the estimates from Experiments II and III, respectively, yielding qualitatively similar results as for the second simulation. While filter bandwidth had no effect on the first simulation (i.e., when the CR was fixed at a high value), in the third simulation filter bandwidth had a small but consistent effect in increasing the modulation power by about a factor of 1.2 when the CR was close to 1 (i.e., in case of a large loss of compression: diamond, star, left-pointing triangle), as a consequence of more harmonic components passing through the filter.

Thus, these results demonstrate that the modulation power of the RP complex tones was low (only slightly above 1, which would imply a flat envelope) and almost independent of both filter bandwidth (top left panel in Fig. 2.7) and compression (middle left panel in Fig. 2.7). In contrast, the modulation power of the SP complex tone increased with increasing loss of compression (almost perfectly linear increase, middle left panel) and, to a minor extent, when increasing filter bandwidth (only at CRs close to 1). Thus, the envelope peakiness of the SP complex tone was increased as compared to the RP envelope up to a factor of 3, mostly as a result of reduced

compression.

This envelope enhancement was estimated as the ratio of the modulation power for the SP complex *vs.* the RP complex (Pmod ratio). The obtained Pmod ratio was then compared with the behavioral F_0 DL ratio for the unresolved conditions (right panels of Fig. 2.7). While no correlation was found when only filter bandwidth was varied (top right panel: $R^2 = 0.19$, $p = 0.24$), a significant correlation was obtained when the individually adjusted loss of compression was introduced to the model (middle right panel: $R^2 = 0.57$, $p = 0.019$). Additionally, adjusting the filter bandwidth did not increase the correlation significantly (bottom right panel: $R^2 = 0.58$, $p = 0.016$). Thus, the modeling outcomes suggested that loss of compression was the dominant factor in enhancing the envelope peakiness of an unresolved SP complex tone relative to its RP counterpart. This enhancement was significantly correlated with the benefit in pitch-discrimination performance for the SP relative to the RP condition.

2.4 Discussion

2.4.1 Relation between behavioral results and envelope representation

The hypothesis of the current study was that if the envelope representation is enhanced for listeners with SNHL (Kale and Heinz, 2010; Henry et al., 2014), pitch cues for unresolved complex tones should also be enhanced if one assumes an envelope coding mechanism for pitch extraction of unresolved harmonics. The pitch-discrimination thresholds measured in the present study (Experiment I) revealed that the HI listeners performed worse than the NH listeners for the RP unresolved conditions (grey-shaded area on middle panels in Fig. 2.2). However, the performance of the HI listeners was similar to that of the NH listeners when the harmonics were added in SP (grey-shaded area on left panels in Fig. 2.2). This finding is in agreement with previous studies showing similar performance of the HI and NH listeners for pitch-discrimination of complex tones with unresolved harmonics (Arehart, 1994; Bernstein and Oxenham, 2006b) and with stronger phase

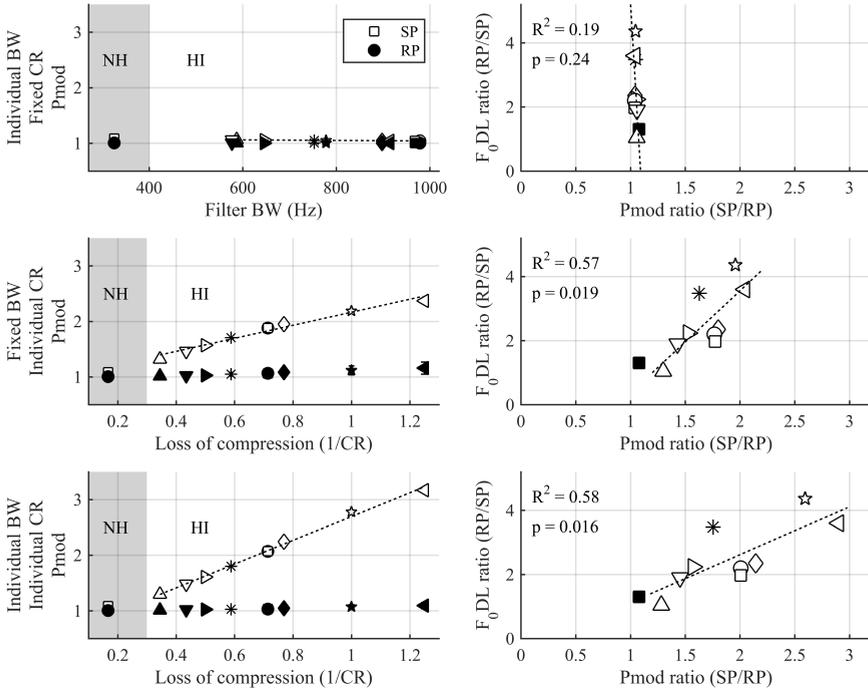


Figure 2.7: Left panels: Envelope modulation power of a complex tone ($F_0 = 100$ Hz) with unresolved harmonics added either in sine phase (SP, open symbols) or in random phase (RP, closed symbols) at the output of a simplified peripheral model. Error bars for the RP condition depict the standard deviations across the 100 iterations. Top panels: the gammatone filter bandwidth was varied according to the individually estimated filter bandwidths from Experiment II, while normal-hearing compression was applied ($CR = 6$); middle panels: cochlear compression was varied according to the estimates from Experiment III, while filter bandwidth was fixed at 325 Hz; bottom panels: both cochlear compression and filter bandwidth were varied. Right panels: correlations between the modulation power ratio (SP/RP) and the behavioral results of Experiment I (mean F_0 DL ratio for the unresolved conditions). The correlations were carried out only across the HI data.

effects for the HI than for the NH listeners (e.g., Moore and Peters, 1992; Moore and Carlyon, 2005; Bernstein and Oxenham, 2006b). In fact, in the presence of a peaky envelope (SP condition), the pitch-discrimination performance of NH listeners increased, on average, by a factor of 1.3 relative to the RP condition (for the two unresolved conditions), while the performance of the HI listeners increased, on average, by a factor of 2.6. Thus, although the overall performance of the HI listeners was not better than that of the NH listeners, these findings suggest that HI listeners benefited more from a peaky signal relative to a signal with a flatter envelope in terms of pitch discrimination than NH listeners did. Hence, the behavioral findings of Experiment I do not rule out an enhanced envelope representation following SNHL. In fact, an envelope enhancement at the output of peripheral stages of the auditory system might be counteracted by other factors limiting the behavioral performance of the HI listeners (e.g., disrupted temporal fine-structure cues, degradation of auditory-nerve coding, higher internal noise level, age-related cognitive deficits). In agreement with this hypothesis, the results of Experiment II revealed significantly lower (better) modulation detection thresholds for the HI listeners (up to 100 Hz) as compared to NH listeners, consistent with previous findings (Moore et al., 1996; Moore and Glasberg, 2001). Thus, when amplitude-modulation detection is based on temporal envelope cues (i.e., when the sidebands are not resolved), the HI listeners showed a higher sensitivity in detecting amplitude modulations imposed on a sinusoidal carrier as compared to NH listeners.

While the larger benefit of HI listeners in pitch-discrimination performance for the SP relative to the RP condition might be a consequence of more harmonics being processed within broader than normal auditory filters, the lower thresholds obtained in Experiment II for HI listeners cannot be explained by the larger number of harmonics within the same auditory filter. In fact, since the sinusoidally amplitude-modulated tones of Experiment II contained only three frequency components (f_c - f_m , f_c , f_c + f_m), broader than normal auditory filters would not lead to additional frequency components passing through the filter. At the very least, the behavioral findings from Experiment I and II suggest that changes that cannot be solely explained by broader auditory filters occurred in the internal envelope representation of listeners with SNHL.

2.4.2 F_0 DL ratio and individual measures of cochlear compression and filter bandwidth

In order to quantify the changes in the internal envelope representation, the increase in pitch-discrimination performance for the SP condition relative to the RP condition (F_0 DL ratio) was considered as an indicator of envelope coding independent of musical abilities and other individual factors. Nine out of 10 HI listeners exhibited F_0 DL ratios larger than those observed in the NH listeners for the two unresolved conditions (at F_0 s of 100 and 125 Hz). An increase of the F_0 DL ratio alone does not necessarily imply an enhancement of envelope coding following SNHL. The larger ratio represents a difference in the salience of temporal envelope cues between the SP and RP complexes, but whether this is the result of an enhancement of envelope cues in the SP condition or a worsening of envelope cues in the RP condition cannot be clarified solely based on the behavioral data. The correlations between the F_0 DL ratios and the individual estimates of cochlear compression and filter bandwidth (Fig. 2.6) revealed a significant correlation for the HI listeners between the increase in the F_0 DL ratio and the reduction of cochlear compression, while no correlation was found with auditory-filter bandwidth. Thus, reducing cochlear compression could account for the increase in performance for the SP condition relative to the RP condition in listeners with SNHL.

Figure 2.8 depicts the correlation between the estimates of auditory-filter bandwidth and cochlear compression obtained from Experiment II and III, respectively. Although not significant, there was a trend of increasing bandwidth with increasing loss of compression ($R^2 = 0.44$, $p = 0.053$). Cochlear compression and auditory-filter bandwidth were found to be physiologically linked and dependent on the cochlear active mechanisms (Ruggero, 1992). Indeed, an earlier psychoacoustic study (Moore et al., 1999) found a significant correlation between filter bandwidth, using the notched-noise method (e.g., Patterson, 1976), and estimated compression, using the growth-of-masking method (Oxenham and Plack, 1997). Thus, the lack of a significant correlation between the two estimates might be due to the use of AM-detection as a measure of frequency selectivity. This is discussed in more detail further below.

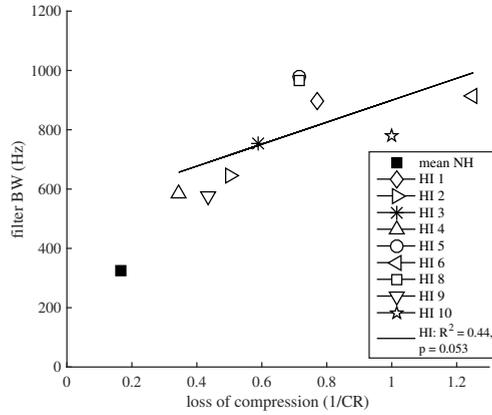


Figure 2.8: Correlation between the estimated auditory-filter bandwidth and loss of cochlear compression across the nine HI listeners (same open symbols as in Fig. 2.1) that participated in both Experiment II and III.

2.4.3 Modeling results and envelope enhancement

Although auditory-filter bandwidth and cochlear compression are physiologically linked, they may have different effects on the envelope at the output of the auditory filters. Therefore, a simplified peripheral model that considers auditory-filter bandwidth and cochlear compression as independent factors was used to qualitatively describe the relative effect of one factor versus the other on the envelope representation of the unresolved complex tones.

The modulation power of a complex tone at the output of the model was used as an indicator of the salience of temporal envelope cues for pitch discrimination of unresolved complexes. The assumption was that the higher the modulation power (i.e., the peakier the envelope), the larger was the salience of temporal pitch cues. Thus, a higher modulation power would correspond to an improved performance in pitch discrimination (i.e., a lower behavioral threshold). The simulation outcomes revealed that reducing cochlear compression and, to a minor extent, increasing the filter bandwidth led to an increase in the modulation power for the unresolved SP complex tone, with reduction of compression clearly being the

dominant factor (left panels in Figure 2.7). In contrast, the modulation power for the RP complex did not vary with either reducing compression or increasing filter bandwidth. Thus, the modeling outcomes suggest that the envelope cues for a RP complex tone may be similar for HI and NH listeners at the output of peripheral stages of the auditory system (provided that audibility is compensated for). Assuming similar processes for NH and HI listeners after the cochlear stages, and assuming a temporal-envelope pitch coding mechanism for unresolved complex tones, one would predict similar performance for the RP condition in listeners with SNHL as compared to NH listeners. However, the behavioral performance of the HI listeners for the RP condition was, on average, worse than for NH listeners. This finding suggests that other individual factors than outer-hair cell damage might limit the performance of the HI listeners for both SP and RP conditions (e.g., disrupted temporal fine-structure cues, degradation of auditory-nerve coding, internal noise). Thus, a possible enhancement of envelope cues following SNHL cannot be revealed based on a comparison of pitch-discrimination thresholds in HI vs. NH listeners, but rather on a comparison between SP vs. RP thresholds, whereby the RP thresholds represent the baseline condition in each listener.

The ratio between the modulation power (P_{mod} ratio) for the SP condition (i.e., where an increase in the envelope peakiness occurred) and the modulation power for the RP condition (i.e., where no increase occurred) was used as an estimate of temporal envelope coding enhancement. The significance in the correlation between the $F_0\text{DL}$ ratio and the P_{mod} ratio (right panels in Fig. 2.7) suggests that the increase in pitch-discrimination performance for the SP relative to the RP condition (i.e., the $F_0\text{DL}$ ratio) can be accounted for by the enhanced envelope for the SP complex tone as compared to the RP baseline condition. Thus, the modeling outcomes revealed that the larger the peripheral loss of cochlear compression, the larger was the enhancement of temporal cues for the SP condition at the output of peripheral stages of the auditory system.

2.4.4 AM-detection as a measure of frequency selectivity

The lack of correlation between the F_0DL ratio and the estimates of auditory-filter bandwidth (Fig. 2.6) may be related to the use of an amplitude-modulation detection task to estimate frequency selectivity. Auditory-filter bandwidth was estimated as the fm where the detection of sidebands was the dominant cue. Thus, the threshold at this fm was probably determined by the upper slope of a filter centered near the lower sideband (Sek and Moore, 1994; Kohlrausch et al., 2000), which in some cases was remote from the center frequency of the stimulus. Although these estimates may not provide a direct measure of frequency selectivity at 2 kHz, but possibly at a lower frequency, they were consistent with the estimates of cochlear compression at 2 kHz (see Figure 2.8). In fact, although not significant, there was a trend of increasing bandwidth with increasing loss of compression. The lack of correlation between the F_0DL ratios and the estimates of auditory-filter bandwidth was, nevertheless, supported by the simulation outcomes, where no or little effect of bandwidth on the envelope peakiness was observed for the SP condition (see Figure 2.7).

2.5 Conclusion

Overall, the results of the pitch-discrimination experiment revealed that the performance of the HI listeners was, on average, similar to that of the NH listeners for the SP unresolved complex tones, and worse for the RP complexes. Thus, the increase in performance for the SP condition relative to its RP counterpart (F_0DL ratio) was significantly larger in the HI listeners as compared to the NH listeners, indicating larger benefits in the presence of a peaky envelope (i.e., the SP condition). This benefit was significantly correlated with the decrease in residual cochlear compression estimated in the same HI listeners. Moreover, the outcomes of a simplified peripheral model revealed that loss of cochlear compression was the dominant factor in enhancing the envelope peakiness of the SP, but not RP unresolved complex tones. This enhancement in the internal envelope representation of unresolved complex tones with harmonics added in SP could account

for the increase in pitch-discrimination performance for the SP relative to the RP condition in listeners with SNHL. Overall, the behavioral results of the present study, together with the modeling outcomes, suggest that listeners with SNHL may have enhanced temporal envelope cues at the output of peripheral stages of the auditory system, primarily as a consequence of a reduced cochlear compression.

Acknowledgments

This work was supported by the Technical University of Denmark. The authors would like to thank Andrew Oxenham and two anonymous reviewers for their constructive feedback.

3

Objective correlates of pitch salience using pupillometry^a

Abstract

Physiological correlates of pitch salience have been investigated in several neuroimaging studies, via functional magnetic resonance imaging and electrophysiological measures. In the present study, a novel approach to objectively estimate pitch salience was used. Pupil dilation was measured as an indicator of the required effort in performing a pitch-discrimination task for complex tones of varying pitch salience. It has been shown that cognitive processing demands of the task can be reflected in the pupil response, whereby pupil size increases with increasing processing effort. The hypothesis was that pupil size would increase with decreasing pitch salience indicating an increase of processing effort to perform the task with less salient stimuli. A group of normal-hearing listeners first performed a behavioral pitch-discrimination experiment, where fundamental frequency difference limens (F_0 DLs) were estimated as a function of F_0 . The obtained pitch-discrimination thresholds suggested that the pitch salience of complex tones filtered in a high spectral region (1.5-3.5 kHz) increased with increasing F_0 . In a second experiment, pupil dilations were recorded while listeners were asked to perform a similar pitch-discrimination

^a This chapter is based on Bianchi, E., Santurette, S., Wendt, D., Dau, T. (2014), *Proc. of Forum Acusticum*, with additional content from Bianchi, E., Santurette, S., Wendt, D., Dau, T. (2015), *Association for Research in Otolaryngology, 38th Mid-Winter Meeting*.

task. Both the stimulus pitch salience and the difficulty of the discrimination task were varied. Although the aim of this study was not to clarify differences in processing effort between musically trained and non-musically trained listeners, the pupil responses showed different trends for musicians and non-musicians. Pupil responses for the musically trained listeners showed the expected trend, whereby pupil size increased with decreasing pitch salience and increasing task difficulty. Non-musically trained listeners also showed an increase in pupil size from the most salient condition to the medium salient condition, while a decrease in pupil size occurred for the least salient condition, probably due to a too demanding pitch-discrimination task. Overall, the findings of this study suggest that pupil responses may reflect both the salience of the pitch-evoking stimuli and the difficulty of the pitch-discrimination task.

3.1 Introduction

The perceptual pitch strength, or pitch salience, of complex tones has been largely addressed in previous studies both behaviorally via pitch-discrimination thresholds (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b; Moore et al., 2006a; Micheyl et al., 2010) and objectively, via either functional magnetic resonance imaging (Penagos et al., 2004; Hall and Plack, 2009; Puschmann et al., 2010; Barker et al., 2011; Norman-Haignere et al., 2013) or electrophysiological measures (e.g., Bidelman and Krishnan, 2009; Gockel et al., 2011; Krishnan et al., 2012). Objective investigations of pitch coding in the human cortical and subcortical structures focused on clarifying the existence of a pitch center that would consistently respond to different pitch-evoking stimuli, with responses proportional to the pitch salience of the stimulus (Hall and Plack, 2009). Using functional magnetic resonance imaging, some studies (Penagos et al., 2004; Norman-Haignere et al., 2013) observed a covariation of neural activity with pitch

salience, whereby resolved complex tones (with a salient pitch) were found to elicit a stronger neural activation than unresolved complex tones (with a less salient pitch) in normal-hearing listeners.

In the present study, a new approach was used to indirectly investigate pitch salience via task-evoked pupil response. Pupil dilation was used as an indicator of the required processing effort in performing a pitch-discrimination task. Since it has been shown that cognitive processing demands of the task can be reflected in the pupil response (see e.g., Janisse, 1977; Beatty, 1982), the hypothesis of the current study was that pupil size would increase with decreasing salience of the stimuli. The aim was to clarify how effort varied during a pitch-discrimination task, depending on the nature of the stimuli (i.e., resolved vs unresolved complex tones) and the salience of the evoked pitch percept. In particular, this study addressed two main questions: whether it is possible to measure a change in processing effort during a pitch-discrimination task via task-evoked pupil response and if this change can be related to the stimulus pitch salience. Pupillometry has so far been used to estimate arousal, focussed attention, memory and cognitive effort mostly in relation to speech processing (Zekveld et al., 2010; Zekveld et al., 2011; Koelewijn et al., 2014), digit-list recall (Granholm et al., 1996; Piquado et al., 2010), mathematical problem solving (Hess and Polt, 1964), or visual tasks (Bradshaw, 1967; Porter et al., 2007; Naber and Nakayama, 2013; Blaser et al., 2014). The novelty of the current study lies in that pupillometry was used here in relation to the processing effort involved in performing a pitch-discrimination task. Although one previous study investigated pupil dilation during a pitch-discrimination task (Kahneman and Beatty, 1967), to the knowledge of the authors this is the first study that used task-evoked pupil dilation to systematically investigate the processing effort in a pitch-discrimination task as an estimate of pitch salience.

Two experiments were performed. A first behavioral pitch-discrimination experiment aimed at estimating the pitch salience of complex tones as a function of the fundamental frequency (F_0). Using a similar paradigm as Bernstein and Oxenham (2006a), difference limens for F_0 (F_0 DLs) were measured for resolved and unresolved complex tones, whereby resolvability was defined by both the frequency range in which the stimuli were filtered and the spacing between the

harmonics (i.e., F_0). In a second experiment, pupil dilation was measured during a similar pitch-discrimination task. The difficulty of the task was individually adjusted according to the results of the first experiment. Pupil size was recorded for conditions below, at and above the individual pitch-discrimination thresholds for resolved (high salience) and unresolved (low salience) complex tones to clarify whether the processing effort to perform a pitch-discrimination task changes when both stimulus salience and task difficulty are varied.

3.2 Experiment I: Difference limens for fundamental frequency

3.2.1 Method

The pitch salience of complex tones was estimated behaviorally with a pitch-discrimination task, where difference limens for F_0 (F_0 DL) were measured as a function of F_0 .

Procedure

A three alternative forced choice (3 AFC) paradigm was used in combination with a weighted up-down method (Kaernbach, 1991) to measure the 75% point on the psychometric function. For each trial, two intervals contained a reference complex tone with a fixed F_0 ($F_{0,\text{ref}}$: 100, 125, 150, 175, 200, 250, 300, 400, 500 Hz) and one interval contained a deviant complex tone with a larger F_0 ($F_{0,\text{ct}}$). The listeners were asked to listen to the stimuli and identify the deviant tone with the highest pitch. The initial difference in F_0 between reference and deviant, ΔF_0 , $(F_{0,\text{ct}} - F_{0,\text{ref}})/F_{0,\text{ref}}$, was set to 20% and was then logarithmically decreased by a varying step size every second reversal. For each run, $F_{0,\text{ref}}$ was roved from trial to trial from a $\pm 5\%$ uniform distribution around the nominal value. The threshold for each run was obtained as the geometric mean of the last 6 reversals. Each participant performed six repetitions of the experiment, of which the first three were considered as training. The final value of F_0 DL was calculated from the mean of the last three repetitions.

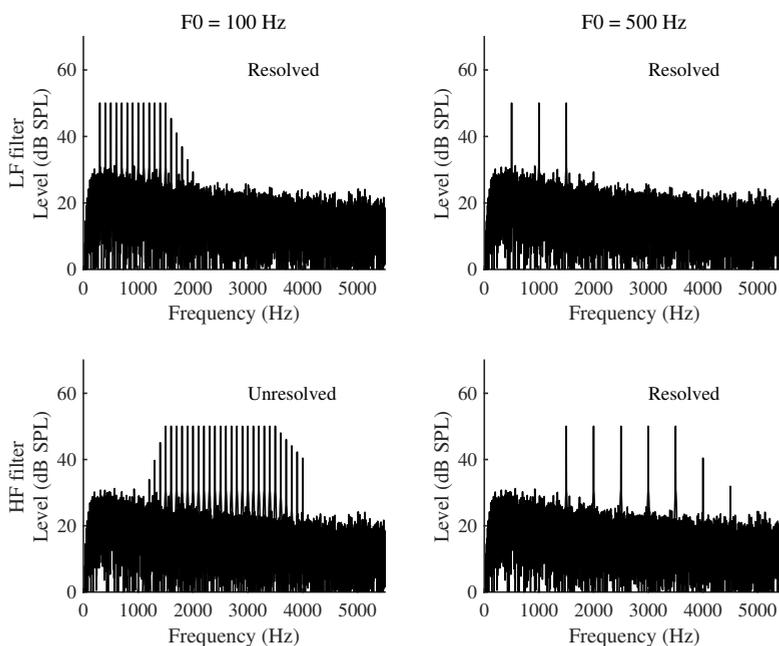


Figure 3.1: Example of stimuli used in Experiment I and II: complex tones with an F_0 of either 100 Hz or 500 Hz, filtered in either a low-frequency region (LF: 0.3-1.5 kHz) or high-frequency region (HF: 1.5-3.5 kHz).

The experiment took place in a double-walled soundproof booth.

Listeners

Fourteen normal-hearing listeners (six females), aged from 22 to 28 years old, participated in the behavioral experiment. Six listeners were musically trained and had played an instrument for more than three years. All listeners had audibility thresholds of less than 20 dB hearing level (HL) at all audiometric frequencies between 125 Hz and 8 kHz.

Stimuli

All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 300-ms complex tones embedded in broadband threshold equalizing noise (TEN, Moore et al., 2000). The level of the TEN was set to 55 dB SPL per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) to mask the combination tones. The complex tones were created by summing harmonic components in sine phase and were filtered in a low-frequency (LF, 300-1500 Hz) or high-frequency (HF, 1500-3500 Hz) region, with 50 dB/octave slopes. Figure 3.1 depicts the two extreme cases of complex tones with an F_0 of either 100 Hz (left panels) or 500 Hz (right panels), filtered either in the LF region (top panels) or HF region (bottom panels) and embedded in TEN. In order to keep the sensation level (SL) of the complex tones approximately constant across listeners, pure tone detection in TEN background was performed at 1.5, 2, and 3 kHz (three repetitions per frequency). For each listener, the mean detection threshold was calculated across the three frequencies and the level of each harmonic component of the complex tone was set at 12.5 dB SL *re* the mean threshold. The sound stimuli were delivered through headphones (Sennheiser HDA 200).

3.2.2 Results

The mean pitch-discrimination thresholds for the 14 listeners are presented in Fig. 3.2. In agreement with previous results (e.g., Bernstein and Oxenham, 2006a), discrimination thresholds for complex tones filtered in a HF-region (filled circles) decreased with increasing F_0 . For small F_0 s (100 and 125 Hz), the listeners needed, on average, a ΔF_0 of about 5.8% to discriminate between reference and deviant tones. With increasing F_0 , F_0 DLs decreased until a baseline value of 1.5% was reached at 400 Hz. The decrease in pitch-discrimination thresholds is assumed to reflect an increase in the resolvability of the harmonics (Bernstein and Oxenham, 2006b), as well as an increase in the availability of temporal fine structure cues (Moore et al., 2006a). Perceptually, the decrease of F_0 DLs reflects an increase in the salience of the evoked pitch percept (Micheyl et al., 2010). A mixed-model ANOVA with F_0 as fixed factor and listeners as random factor was fit to the HF

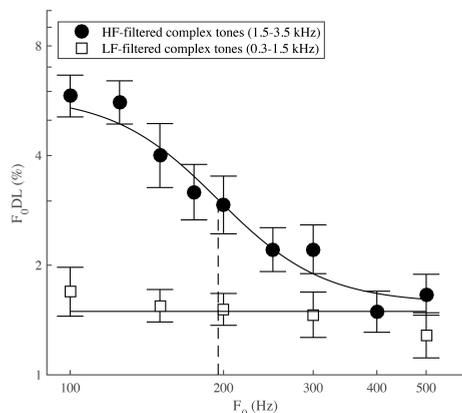


Figure 3.2: Mean pitch-discrimination thresholds for 14 listeners, for complex tones filtered in a low-frequency (LF; open squares) and high-frequency (HF; filled circles) region. The top curve depicts the fitted sigmoid to the HF data, whereas the lower curve shows the mean LF-threshold. Error bars depict the standard error of the mean.

thresholds and confirmed a significant effect of F_0 ($F(8,125) = 29.6$; $p < 0.0001$). A sigmoid function was fitted to the HF-thresholds and the point halfway between the maximum and the minimum thresholds (on a log scale) occurred at a $\Delta F_{0,50\%}$ of 3% when the transition F_0 s ($F_{0,tr}$) was 195 Hz (dashed line in Fig. 3.2). The $F_{0,tr}$ is thought to reflect the transition point from unresolved to resolved harmonics of the complex tone (Bernstein and Oxenham, 2006b). The mean discrimination thresholds for the LF-filtered complex tones (open squares) were, on average, around 1.5% and did not show a significant effect of F_0 ($F(4,69) = 1.72$; $p = 0.16$).

3.3 Experiment II: Pupillometry

3.3.1 Method

In the second experiment, pupil dilation was measured during a pitch-discrimination task. Pupil size was recorded for conditions of varying pitch salience, below, at and above the individual pitch-discrimination threshold to investigate how

processing effort varied with stimulus salience and task difficulty.

Procedure

The listeners were presented with three consecutive complex tones, two references with a fixed F_0 and one deviant with a higher F_0 (see Fig. 3.3(c)). For each listener, the difference in F_0 between reference and deviant, ΔF_0 , was adjusted according to the thresholds of Experiment I (cf. Sec. 3.3.1). Each trial consisted of 2 seconds of initial silence, followed by 3.8 seconds of sound stimulation. Figure 3.3(c) depicts the sound stimulation paradigm, consisting of 2.3 seconds of initial baseline (TEN at 55 dB/ERB), followed by 1.5 seconds of stimulation with complex tones (two references and one deviant) embedded in TEN. After stimulus presentation, the listeners had 3 seconds to identify the deviant by pressing a key on the keyboard. The listeners were asked to fixate on a marker on the screen and blink as little as possible while performing the task. Pupil size was recorded for the whole duration of each trial (i.e., 2 s-silence, 3.8 s-stimulation, 3 s-task) using an eye-tracking device (EyeLink 1000, SR Research Ltd), which used infrared tracking technology to measure the pupil area (in arbitrary units). Percentage of correct deviant identification was also measured for each condition. The experiment took place in a double-walled soundproof booth. After a short training session, each listener performed 15 repetitions of each stimulus condition (i.e., 90 trials), for a total duration of the experiment of 20 minutes.

Listeners

Eleven listeners (7 females) participated in Experiment II. Six of these listeners had participated previously in Experiment I. Six listeners were musically trained (indicated with asterisks in Table 3.1) with at least three years of formal musical training and five listeners had no prior musical experience.

Stimuli

In order to keep the difficulty of the task similar across listeners, the difference in F_0 between reference and deviant (ΔF_0) was set at the point halfway between

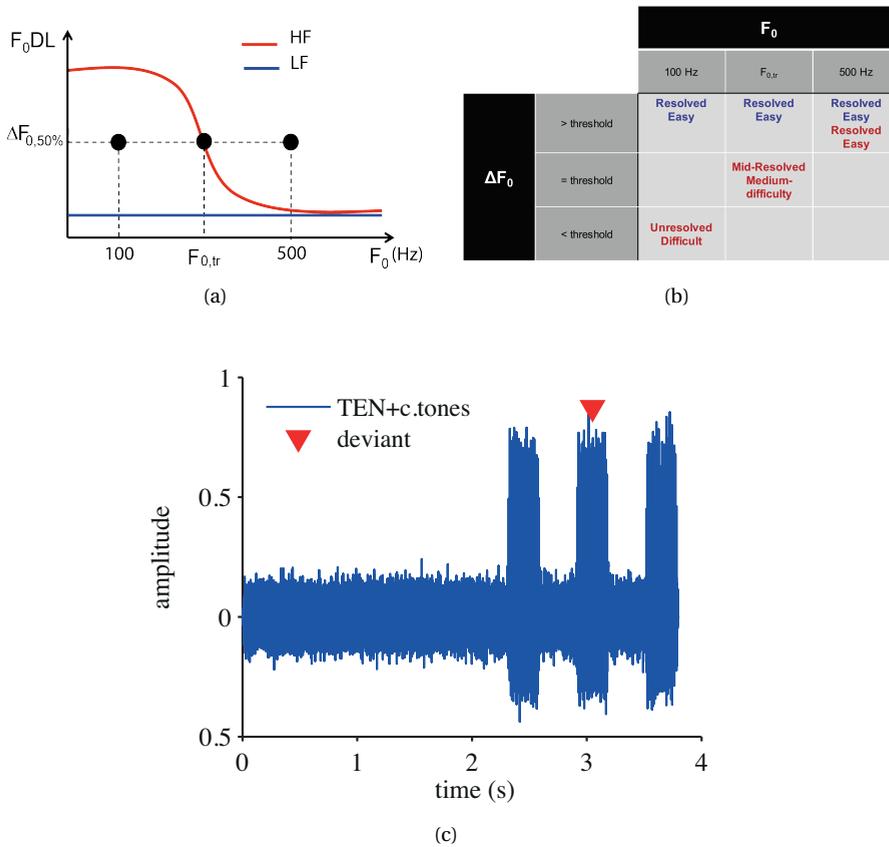


Figure 3.3: a) Schematic illustration of the stimulus conditions used in Experiment II. Pitch-discrimination was performed for complex tones at three different F_0 s (100 Hz, $F_{0,tr}$ and 500 Hz). The difference in F_0 between the reference and deviant tone was adjusted at an individually fixed value of $\Delta F_{0,50\%}$ (cf. Table 3.1). Both frequency regions (LF, blue curve; HF, red curve) were tested at the same $\Delta F_{0,50\%}$, yielding conditions of varying task difficulty (cf. Fig. 3.3(b)). b) Summary of the six conditions presented in Experiment II. Complex tones filtered in a LF region are depicted in blue and complex tones filtered in a HF region are depicted in red. Resolvability and, thus, pitch salience increase with increasing F_0 for the HF-filtered complex tones, while all LF conditions are resolved independent of F_0 . Task difficulty was varied by adjusting ΔF_0 to be below (high difficulty), at (medium difficulty) and above (low difficulty) the individual pitch-discrimination thresholds from Experiment I. c) Stimulus presentation for Experiment II. Each trial consisted of 2.3 s of baseline with noise (TEN at 55 dB/ERB), followed by three complex tones embedded in TEN for a total of 3.8 s of sound stimulation. The downwards pointing triangle indicates the deviant complex tone with a higher F_0 than the two references. The listeners' task was to identify the deviant by pressing a key on the keyboard after stimulus presentation. Pupil responses were recorded for the whole duration of the trial.

maximum and minimum of the sigmoid fitted to the individual F_0 DLs ($\Delta F_{0,50\%}$, see Fig. 3.3(a)). For the listeners that did not participate in Experiment I, ΔF_0 was set at 3%, i.e., at the mean $\Delta F_{0,50\%}$ obtained for all 14 listeners (see Fig. 3.2). Table 3.1 depicts the values of $\Delta F_{0,50\%}$, as well as the corresponding $F_{0,tr}$ s, which were used in Experiment II for each listener.

Similar complex tones as for the behavioral experiment (see Sec. 3.2.1) were used in the current experiment. Overall, six conditions were tested, namely, complex tones at three different F_0 s (100 Hz, $F_{0,tr}$, and 500 Hz), filtered in either a low- or high-frequency regions (see Fig. 3.3(a) and 3.3(b)). The three LF-conditions were tested at a ΔF_0 above the mean pitch-discrimination threshold, yielding to three high-salience conditions tested with an easy discrimination task. The three HF-conditions were tested below (100 Hz), at ($F_{0,tr}$), and above the pitch-discrimination threshold (500 Hz), yielding to conditions with concomitantly increasing pitch salience and decreasing task difficulty (summarized in Fig. 3.3(b)).

Data Analysis

For each trial, the mean baseline was calculated by averaging the mean pupil size in the 0.7-s interval preceding the beginning of stimulation with complex tones. The mean baseline was then subtracted from each trial. The mean pupil size across the 15 repetitions was calculated for each condition, and pupil sizes exceeding ± 3 standard deviations from the mean value were coded as eye blinks. Trials containing more than 15% of samples as eye blinks during complex-tone stimulation were excluded from the analysis (Zekveld and Kramer, 2014; Wendt et al., 2016). The data were filtered by a 15-point moving average smoothing filter. All statistical analyses were performed in MATLAB.

3.3.2 Results

Although the aim of this study was not to clarify differences in processing effort between musically trained and non-musically trained listeners, the pupil responses showed different trends for the two groups of listeners. Thus, results for musicians and non-musicians are presented and discussed separately. The mean pupil

Table 3.1: Individual values of $\Delta F_{0,50\%}$ and $F_{0,tr}$ used for the pupillometry experiment. Asterisks denote the musically trained listeners.

Listener	$\Delta F_{0,50\%}$ [%]	$F_{0,tr}$ [Hz]
1*	3.8	218.6
2*	2.5	215.8
3*	2	142
4*	3	200
5*	3	200
6*	3	200
7	3	200
8	3	200
9	3.3	252.9
10	3.6	213.9
11	6	189.2

dilations for the six musically trained listeners are presented in Fig. 3.4 (top panels), and the mean dilations for the five non-musically trained listeners in Fig. 3.4 (bottom panels). The left panels in Fig. 3.4 depict the mean pupil responses for each group of listeners as a function of time. The time axis refers to the beginning of stimulation with complex tones (i.e., time zero in Fig. 3.4 refers to time 2.3 s in Fig. 3.3(c)). Four conditions are presented: the mean dilation for the three LF control conditions (gray curve) and dilations for each of the three HF-conditions, respectively, for an F_0 of 100 Hz (black curve), $F_{0,tr}$ (red curve), and 500 Hz (blue curve). For all conditions, pupil size increased during stimulation with complex tones until reaching a maximum dilation point at about 2 seconds after stimulus (complex tones) onset. After the maximum dilation point, pupil size decreased at different rates depending on the condition tested. For the musically trained listeners, the least salient condition (black curve) led to a longer-sustained pupil dilation, while the two most salient conditions (gray and blue curves) led to a faster decay. For the non-musically trained listeners, the medium-salient condition (red curve) led to the largest dilations both during and after stimulation. The time-averaged pupil size was calculated from the maximum dilation point (1.9 s, dashed line in Fig. 3.4) until 3.5 seconds after stimulus (complex tones) onset. The ob-

tained time-averaged values of pupil dilation, as well as the percentage of correct deviant identifications are presented for each condition in the right panels of Fig. 3.4. Although a mixed-model ANOVA with conditions as fixed factor and listeners as random factor did not reveal a significant general effect of conditions on the time-averaged pupil dilation (musicians: $F(3,23) = 3.2$, $p = 0.053$; non-musicians: $F(3,19) = 3.15$, $p = 0.065$), a t-test revealed a significant difference in pupil size between the least salient condition (black bar) and the most salient condition (blue bar) for the musicians ($p = 0.035$). It is, however, unclear whether the increase in pupil size was driven by the decrease in the salience of the stimuli (from resolved to unresolved harmonics) or by the increase in the task difficulty (from a condition tested above the pitch-discrimination threshold to a condition tested below the threshold). For the non-musicians, pupil size increased from the 500 Hz- to the $F_{0,tr}$ -condition, but it did not increase further for the 100-Hz condition. This finding, together with the very low performance of the non-musicians in the 100-Hz condition (42% correct performance), suggests that the task for this condition was probably too demanding for the non-musically trained listeners. Effort might have dropped as a consequence of a task exceeding the participant's ability (Zekveld and Kramer, 2014).

3.4 Discussion

The pitch salience of complex tones was first estimated behaviorally via F_0 DLs. In agreement with previous findings (e.g., Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b), a significant decrease of F_0 DLs was obtained when increasing the F_0 of complex tones filtered in a HF-region (Fig. 3.2, filled circles). This decrease is thought to reflect the transition point from unresolved to resolved harmonics (Bernstein and Oxenham, 2006b) as well as the increase in the availability of temporal fine structure cues (Moore et al., 2006a), when decreasing the lowest harmonic number present in a complex tone. Additionally, a decrease in the F_0 DLs reflects an increase in the salience of the evoked pitch percept (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Micheyl et al., 2010). As a

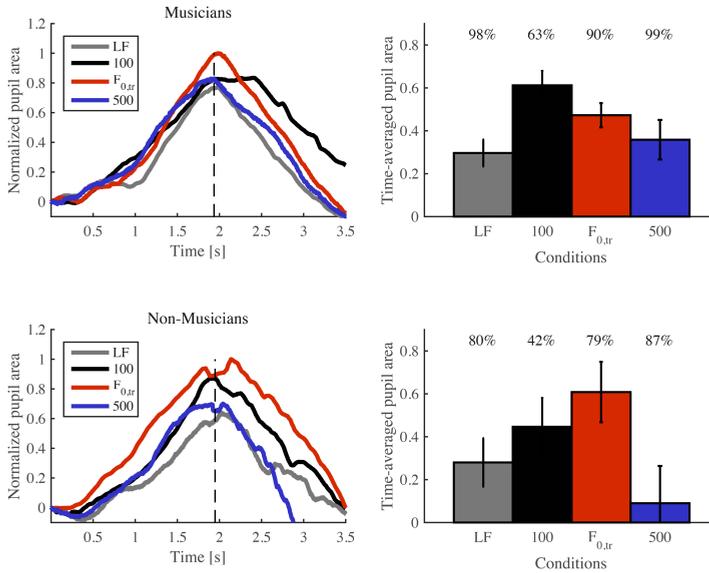


Figure 3.4: Left panels: mean pupil dilation across listeners as a function of time. The time-axis refers to the beginning of stimulation with complex tones. Mean results averaged across the three LF-conditions are depicted in gray; results for the HF-conditions at F_0 s of 100 Hz, $F_{0,ir}$ (indicated as 200 Hz in the legend), and 500 Hz are depicted in black, red and blue, respectively. Right panels: time-averaged value of pupil dilation, calculated from the maximum dilation point (around 2 s) until 3.5 s after stimulus onset. Error bars depict the standard error of the mean at the maximum pupil dilation point. The values on top of the right panels indicate the percentage of correct deviant identification in each condition. Top panels: Mean results for six musically trained listeners. Bottom panels: Mean results for five non-musically trained listeners.

control condition, pitch salience was also estimated for complex tones filtered in a LF-region. The results showed that increasing F_0 did not have any significant effect on the pitch-discrimination thresholds for the LF conditions (Fig. 3.2, open squares). This is consistent with narrower auditory filters at low frequencies (on a linear scale, Glasberg and Moore, 1990), which allow the harmonics to be resolved already for small F_0 s.

The aim of this study was to clarify whether pupil dilation, considered as an indicator of the required effort to perform a pitch-discrimination task, could reflect the pitch salience evoked by resolved and unresolved complex tones. The hypothe-

sis was that the processing effort during a pitch-discrimination task would increase with decreasing pitch salience and with increasing task difficulty. Thus, it was predicted that effort would be highest in the low-salience condition tested below threshold (i.e., the 100 Hz-condition in the HF-region) and lowest in the salient conditions tested above threshold (i.e., LF-conditions and 500 Hz-condition in the HF-region). Pupil dilation was, as expected, smaller for the most salient conditions (gray and blue bars in Fig. 3.4) than for the other conditions for both musically trained and non-musically trained listeners. However, only the musically-trained listeners showed the largest pupil dilation (i.e., the highest processing effort) for the least salient condition (black bar). Interestingly, the non-musicians showed the largest processing effort for the medium salient condition (red bar) and a decrease of effort for the least salient condition tested below threshold (black bar). This finding might be explained in the light of previous studies (Pook, 1973; Granholm et al., 1996; Zekveld and Kramer, 2014), where pupil responses were investigated during highly demanding tasks. It was found that pupil size was the largest for medium-difficulty conditions, while it decreased for very difficult conditions (i.e., when the percentage of correct responses in the behavioral task was lower than 50%). Thus, in the presence of a task that exceeded the participant's ability, the decrease in pupil dilation reflected the fact that the listener often "gave up" (cognitive processing overload; Zekveld and Kramer, 2014). In the current study, the difficulty of the task in the below threshold condition (black bar) was higher for the non-musicians than for the musicians. In fact, the behavioral thresholds obtained in Experiment I were larger for the non-musicians than for the musicians (F_0 DLs of 7% vs 4% at 100 Hz). Thus, the $\Delta F_{0,50\%}$ used in Experiment II for the 100-Hz condition probably corresponded to a lower point on the psychometric function for the non-musically trained than for the musically trained listeners, leading to a too difficult task. This was also confirmed by the low score in correct deviant identification (42% for the non-musically trained vs 63% for the musically trained group). Additionally, it should be noted that since the discrimination task appeared not to be equally demanding across participants, the effort related to a specific condition could not be directly compared across the two groups of listeners.

Finally, in the current paradigm, stimulus-evoked pitch salience and the dif-

difficulty of the task (defined by the difference between $\Delta F_{0,50\%}$ and the listener's threshold) concomitantly varied, so it was not possible to disentangle the effect of one factor relative to the other on the pupil responses. Since all salient conditions were tested with an easy task and the least salient condition was only tested with a difficult task, it is unclear whether pupil dilations were driven by stimulus salience per se and/or by task difficulty. Overall, the observed increase in pupil size suggests an increasing effort required for performing the task with either decreasing F_0 for the HF-filtered conditions (i.e., with decreasing harmonic resolvability and pitch salience) or/and with increasing task difficulty.

3.5 Conclusion

The present study addressed two main questions: whether it is possible to measure a change in processing effort during a pitch-discrimination task and if this change can be related to pitch salience. The findings of this study revealed that it is possible to measure processing effort during a pitch-discrimination task via task-evoked pupil response. However, it remains unclear whether this change was driven by pitch salience and/or by task difficulty. When the difficulty of the task was individually adjusted according to the behavioral data, the musically-trained listeners showed pupil dilations in line with the initial hypothesis, whereby processing effort increased with decreasing the salience of the stimuli and increasing the difficulty of the task. The non-musically trained listeners also showed an increase in pupil size from the easy conditions to the medium-difficulty condition, but a decrease in processing effort was obtained for the least salient condition, probably as a result of a cognitive processing overload (i.e., the listeners gave up when the task exceeded their abilities). Future work may clarify the relative importance of stimulus salience and task difficulty for pupil dilations during a pitch-discrimination task.

Acknowledgments

This work was supported by the Technical University of Denmark.

4

Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort^a

Abstract

Musicians typically show enhanced pitch discrimination abilities compared to non-musicians. The present study investigated this perceptual enhancement behaviorally and objectively for resolved and unresolved complex tones to clarify whether the enhanced performance in musicians can be ascribed to increased peripheral frequency selectivity and/or to a different processing effort in performing the task. In a first experiment, pitch discrimination thresholds were obtained for harmonic complex tones with fundamental frequencies (F_0 s) between 100 and 500 Hz, filtered in either a low- or a high-frequency region, leading to variations in the resolvability of audible harmonics. The results showed that pitch discrimination performance in musicians was enhanced for resolved and unresolved complexes to a similar extent. Additionally, the harmonics became resolved at a similar F_0 in musicians and non-musicians, suggesting similar peripheral frequency selectivity in the two groups of listeners. In a follow-up experiment, listeners' pupil dilations were measured as an indicator of the required effort in performing the same pitch discrimination task for conditions

^a This chapter is based on Bianchi, E., Santurette, S., Wendt, D., Dau, T. (2016), *J. Assoc. Res. Otolaryngol.* 17, pp. 69-79.

of varying resolvability and task difficulty. Pupillometry responses indicated a lower processing effort in the musicians versus the non-musicians, although the processing demand imposed by the pitch discrimination task was individually adjusted according to the behavioral thresholds. Overall, these findings indicate that the enhanced pitch discrimination abilities in musicians are unlikely to be related to higher peripheral frequency selectivity and may suggest an enhanced pitch representation at more central stages of the auditory system in musically trained listeners.

4.1 Introduction

Musicians typically show enhanced pitch discrimination ability compared to non-musicians, consistent with the finding that musicians are more sensitive to some acoustic features critical for both speech and music processing (e.g., Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006; Anderson and Kraus, 2011). Although there is evidence of anatomical changes in the musicians' auditory and motor-related structures and enhanced neural responses to sounds (for a review, see Zatorre and Zarate, 2012; Barrett et al., 2013), it is still unclear which mechanisms underlie a perceptual pitch discrimination advantage. A recent study suggested an enhancement of peripheral frequency selectivity in musicians, whereby narrower auditory filters were psychoacoustically estimated in musically trained listeners as compared to non-musicians (Bidelman et al., 2014; Bidelman et al., 2016). Other studies observed an increased subcortical neural synchrony in response to speech in noise resulting in a more precise temporal and spectral representation of the signal (Parbery-Clark et al., 2009; Anderson and Kraus, 2011). It has been suggested that a training-dependent component might be responsible for enhancing neural responses to sounds (e.g., Zatorre and Zarate, 2012; Barrett et al., 2013), although not all studies reporting neural coding enhancements in musicians have shown correlations with the extent of musical training (Parbery-Clark et al., 2009; Parbery-Clark et al., 2012)

To clarify which mechanisms lead to enhanced pitch discrimination performance in musicians, the current study investigated complex-tone pitch discrimination behaviorally and objectively in musicians versus non-musicians. While an enhancement in pitch discrimination was previously reported for pure tones (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001) and complex tones containing resolved harmonics (Micheyl et al., 2006; Allen and Oxenham, 2014), pitch discrimination performance for unresolved complexes in musicians versus non-musicians has not been reported so far. Resolved complex tones contain low-numbered harmonics which are processed by individual auditory filters on the basilar membrane and, thus, convey both frequency and time information. Unresolved complex tones consist of high-numbered harmonics which interact within a given auditory filter and do not convey frequency information about the individual harmonics. As a result, the pitch of resolved complex tones may be retrieved by either spectral and/or temporal cues, whereas the pitch of unresolved complex tones can only be retrieved via temporal coding mechanisms (for a review, see De Cheveigné, 2005). The hypothesis of the current study was that a greater enhancement in performance for resolved (vs. unresolved) complex tones would suggest a finer spectral resolution along the auditory system in musicians. In contrast, a similar enhancement for resolved and unresolved complexes would suggest a greater general ability to attend to and extract pitch-related features following musical training.

Three experiments were performed. First, pitch discrimination thresholds were estimated as a function of the fundamental frequency (F_0) to clarify whether musical training improved discrimination of complex tones containing resolved versus unresolved harmonics to the same extent. Moreover, the transition point at which harmonics became resolved was derived from the individual pitch discrimination thresholds and used as an estimate of auditory filter bandwidths to compare peripheral frequency selectivity in musicians versus non-musicians. This approach to estimate filter bandwidths was suggested by Bernstein and Oxenham, 2006b, who showed a significant correlation between traditional measures of frequency selectivity and the transition point for harmonic resolvability.

Second, pupil responses were recorded as a physiological correlate of process-

ing effort, while the listeners were performing the same pitch discrimination task. The rationale behind this was to investigate how processing effort (as reflected by task-evoked pupil dilations; e.g., Janisse, 1977; Beatty, 1982) varied in musicians and non-musicians, when varying the processing demand imposed by the listening condition. While it has been shown that processing effort increases with increasing the processing demand of the listening condition for speech (Johnsrude and Rodd, 2015), to the knowledge of the authors, this is the first study to investigate pupil dilation during a pitch discrimination task with varying harmonic resolvability and task difficulty. While in a previous study (Bianchi et al., 2014), pupil dilations were measured for conditions with concomitantly varying harmonic resolvability and task difficulty, a new experimental design was used here to disentangle the effects of resolvability and task difficulty on pupil dilations. In experiment 2, pitch discrimination thresholds were measured behaviorally at three F_0 s (i.e., three levels of resolvability) and at three different points of the psychometric function (i.e., three levels of task difficulty). The individual thresholds were then used in the pupillometry measurement (experiment 3) to set conditions that matched in task difficulty and resolvability across listeners. As the processing demand imposed by the pitch discrimination task was, thus, similar for musicians and non-musicians, the hypothesis was that pupil dilations (indicating required processing effort to perform the task) should be similar in the two groups of listeners, if one assumes similar pitch representations along the auditory pathway in musicians and non-musicians. In contrast, smaller pupil dilations (indicating lower processing effort) in musicians would suggest an enhanced pitch representation along the auditory system following musical training (e.g., finer spectral resolution and/or finer F_0 representation at central stages of the auditory system).

4.2 Method

4.2.1 Experiment 1: Behavioral Pitch Discrimination Thresholds

Pitch discrimination thresholds for complex tones were estimated behaviorally via difference limens for F_0 (F_0 DLs) as a function of F_0 . The aim was to clarify whether

musical training improved pitch discrimination of resolved and unresolved complex tones to the same extent. The resolvability of the complex tones was varied by filtering the stimuli in a high-frequency (HF) region and by systematically varying F_0 , such that neighboring harmonics would become resolved with increasing F_0 . Complex tones filtered in a low-frequency (LF) region were used as a baseline (control) condition, since here the auditory filters are narrower and the stimuli always contain resolved harmonics for the same range of F_0 s.

Listeners

Six musicians (more than 3 years of formal musical training, four females) and eight non-musicians (no formal musical training, two females) participated in experiment 1. Ages ranged from 22 to 28 years, with a mean of 25.3 and a median of 25 years. None of the listeners was a tone language speaker. All participants provided written informed consent to participate in the study. All experiments were approved by the Science Ethics Committee for the Capital Region of Denmark. All listeners had audibility thresholds of less than 20 dB hearing level (HL) at all audiometric frequencies between 125 and 8 kHz. The experiment was carried out in a double-walled soundproof booth. The listeners were asked to listen to the stimuli and identify the complex tones with the highest pitch by pressing a response button on the keyboard.

Stimuli

All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 300-ms complex tones embedded in broadband (20–10 kHz) threshold equalizing noise (TEN, Moore et al., 2000). The stimuli were delivered monaurally to the right ear through headphones (Sennheiser HDA 200). The sound pressure level (SPL) of the TEN was set to 55 dB per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) to mask combination tones. The complex tones were created by summing harmonic components in sine phase and were bandpass-filtered in a LF (300–1500 Hz) or HF (1500–3500 Hz) region with 50 dB/oct. slopes. Fourteen conditions were tested in total (nine F_0 s in the HF region at the F_0 s of 100,

125, 150, 175, 200, 250, 300, 400, and 500 Hz; five conditions in the LF region at the F_0 s of 100, 150, 200, 300, and 500 Hz). In order to keep the sensation level (SL) of the complex tones approximately constant across listeners, pure-tone detection in a TEN background was performed at 1.5, 2, and 3 kHz (three repetitions per frequency) before the experiment. For each listener, the mean detection threshold was calculated across the three frequencies and the level of each component of the complex tone (within the passband) was set to 12.5 dB above the mean threshold.

Procedure

A three-alternative forced-choice (3 AFC) paradigm was used in combination with a weighted up-down method (Kaernbach, 1991) to measure the 75% point on the psychometric function. In each trial, two intervals contained a reference complex tone with a fixed F_0 ($F_{0,\text{ref}}$) and one interval contained a deviant complex tone with a larger F_0 ($F_{0,\text{dev}}$). $F_{0,\text{ref}}$ was roved from trial to trial from a $\pm 5\%$ uniform distribution around the nominal value. For each run, the initial difference in F_0 between reference and deviant, ΔF_0 , $(F_{0,\text{dev}} - F_{0,\text{ref}}) / F_{0,\text{ref}}$, was set to 20% and was then logarithmically decreased by a varying step size every second reversal. The threshold for each run was obtained as the geometric mean of the last six reversals. Each listener performed six repetitions of the experiment, of which the first three were considered as training. The conditions were presented in random order within each repetition. The final value of $F_{0,\text{DL}}$ was calculated from the geometric mean of the last three repetitions.

4.2.2 Experiment 2: Effects of Harmonic Resolvability and Task Difficulty

In experiment 2, $F_{0,\text{DL}}$ s were measured as in experiment 1, for a subset of F_0 s and at three different points on the psychometric function. The aim was to behaviorally determine the individual thresholds for different performance levels, such that task difficulty could be matched across listeners in experiment 3.

		F_0		
		100 Hz	200 Hz	500 Hz
Task difficulty	90%	Unresolved Easy	Mid-Resolved Easy	Resolved Easy
	75%	Resolved Medium-difficulty		Resolved Medium-difficulty
		Unresolved Medium-difficulty	Mid-Resolved Medium-difficulty	Resolved Medium-difficulty
60%	Unresolved Difficult	Mid-Resolved Difficult	Resolved Difficult	

Figure 4.1: Summary of the 11 conditions used in experiments 2 and 3. Complex tones filtered in a LF region are depicted in blue and complex tones filtered in a HF region are depicted in red. Task difficulty was varied by adjusting ΔF_0 according to the individual pitch-discrimination thresholds at the 60% (high difficulty), 75% (medium difficulty) and 90% (low difficulty) points on the psychometric function.

Listeners

Ten musicians (more than 4 years of formal musical training, six females) and 10 non-musicians (no formal musical training, four females) participated in the behavioral experiment. Ages ranged from 23 to 28 years, with a mean of 25.8 and a median of 26 years. All listeners had audibility thresholds of less than 20 dB HL at all audiometric frequencies between 125 and 8 kHz.

Stimuli

The complex tones were generated as in experiment 1. Figure 4.1 shows a summary of the 11 tested conditions (nine conditions in the HF region, 60, 75, and 90% points on the psychometric function at the F_0 s of 100, 200, and 500 Hz; two conditions in the LF region, 75% point at the F_0 s of 100 and 500 Hz).

Prodedure

A similar 3 AFC paradigm as in experiment 1 was used here in combination with a weighted up- down method to track the 60, 75, and 90% points on the psychometric function. Pitch discrimination thresholds were measured at three F_0 s ($F_{0,\text{ref}}$ 100, 200, 500 Hz), corresponding to three levels of resolvability for the HF-filtered complex tones (100 Hz, unresolved components; 200 Hz, transition point; 500 Hz, resolved components). Each listener performed five repetitions of the experiment, of which the first two were considered as training.

4.2.3 Experiment 3: Pupil Dilations During Pitch Discrimination

In experiment 3, pupil dilation was measured during a pitch discrimination task. Pupil size was recorded for the 11 conditions of experiment 2 (see Figure 4.1) to investigate how processing effort varied with resolvability and task difficulty.

Listeners

The same listeners that participated in experiment 2 also performed the pupillometry measurement.

Stimuli

Similar complex tones as for experiment 1 were used in the current experiment. For each listener and condition, the difference in F_0 between reference and deviant, ΔF_0 , was set at the behavioral threshold obtained in experiment 2. Thus, pupil dilations were measured at three task difficulty levels (60% point on the psychometric function, high task difficulty; 75%, medium task difficulty; 90%, low task difficulty), three resolvability levels in the HF region (100 Hz, only unresolved harmonics; 200 Hz, transition point from experiment 1; 500 Hz, resolved harmonics), and two control conditions in the LF region (resolved complexes at medium task difficulty). These two control conditions were chosen to control that pupil responses to the HF stimuli were due to changes in the resolvability of the harmonics and not to changes in F_0 .

Procedure and equipment

The listeners were presented with three consecutive complex tones, two references with a fixed F_0 and one deviant with a higher F_0 . The deviant was presented in a random position among the references (either as first, second, or third stimulus). Each trial consisted of 2 s of initial silence, followed by 3.8 s of sound stimulation. Sound stimulation comprised 2.3 s of initial baseline (TEN at 55 dB/ERB), followed by 1.5 s of stimulation with complex tones embedded in TEN (two references and one deviant). After stimulus presentation, the listeners had 3 s to identify the deviant by pressing a key on the keyboard. During the whole duration of the trial (8.8 s), listeners were asked to fixate a dot that was presented on the computer screen, while an eye tracker system (EyeLink 1000 Plus, SR Research Ltd) was used with a sampling rate of 1000 Hz to monitor the participants' pupil area. The visual stimulus was presented on a 22 inches computer screen with a resolution of 1680×1050 pixels. Participants were seated 60 cm from the computer screen, and a chin rest was used to stabilize their head. The eye tracker sampled only from the left eye. The listeners' task was to identify the complex tones with the highest pitch. The percentage of correct deviant identification was also measured for each condition. After a short training session, each listener performed 15 repetitions of each stimulus condition (i.e., 165 trials), randomly presented, for a total duration of the experiment of 40 min.

Data Analysis

For each trial, the mean baseline was calculated by averaging the mean pupil size in the 0.7-s interval preceding the beginning of stimulation with complex tones. The mean baseline was then subtracted from each trial. The mean pupil size across the 15 repetitions was calculated for each condition, and pupil sizes exceeding ± 3 standard deviations from the mean value were coded as eye blinks. Trials containing more than 15% of samples as eye blinks during complex-tone stimulation were excluded from the analysis (Zekveld and Kramer, 2014). To avoid artifacts, samples in a range from 35 to 70 ms around eye blinks were discarded from the analysis. The data were filtered by a 15-point moving average smoothing

filter. All statistical analyses were performed in MATLAB.

4.3 Results

4.3.1 Behavioral Pitch Discrimination Thresholds

Figure 4.2 depicts the mean pitch discrimination thresholds obtained in experiment 1 for six musicians (left panel) and eight non-musicians (right panel). The thresholds for both groups of listeners showed similar trends, whereby F_0 DLs for the HF-filtered complex tones (filled circles in Fig. 4.2) decreased with increasing F_0 , whereas they were independent of F_0 for the LF-filtered complex tones (open squares in Fig. 4.2). Thresholds for non-musicians were, on average, larger than thresholds for musicians by a factor of 1.72. All resolved conditions (LF conditions and HF conditions for F_0 s larger than the transition point, $F_{0,tr}$) were larger by a factor of 1.76 and all unresolved conditions (HF conditions for F_0 s smaller than $F_{0,tr}$) by a factor of 1.61.

A mixed model with group and F_0 as main effects and listeners as random factor nested in group was fit to the set of data, for both LF and HF results. The analysis confirmed a significant group effect for both the HF-filtered conditions ($F(1,125) = 5.14$; $P = 0.043$) and the LF-filtered conditions ($F(1,69) = 11.43$; $P = 0.006$), while the interaction factor of group and F_0 was not significant ($F(8,125) = 0.27$; $P = 0.973$ and $F(4,69) = 1.29$; $P = 0.288$), indicating a similar effect of F_0 in the two groups of listeners. Additionally, the analysis revealed a significant effect of F_0 for the HF-filtered conditions ($F(8,125) = 27.62$; $P < 0.0001$) and no significant effect of F_0 for the LF-filtered conditions ($F(4,69) = 1.78$; $P = 0.16$). The current findings for the HF-filtered conditions are in agreement with previously reported pitch discrimination thresholds (Bernstein and Oxenham, 2006b), where the improvement in performance with F_0 was thought to reflect the progressive increase in the resolvability of the harmonics. A sigmoid function was fitted to the mean HF thresholds, and the transition point ($F_{0,tr}$, vertical dashed line in Fig. 4.2) yielding the F_0 DL halfway (on a log scale) between the maximum and minimum values of the fitted sigmoid was used here as an estimate of peripheral

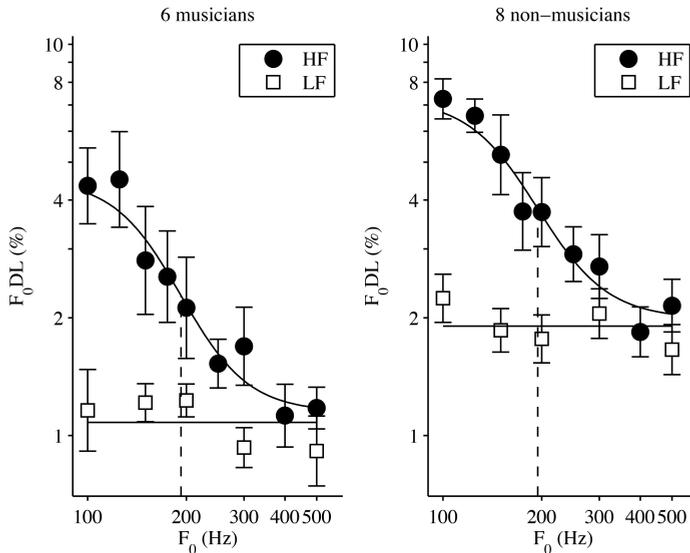


Figure 4.2: Mean pitch discrimination thresholds (F_0DL s) as a function of F_0 , for six musicians (left panel) and eight non-musicians (right panel). The filled circles depict the thresholds (geometric mean) for the high-frequency (HF)-filtered complex tones, while the open squares depict the thresholds (geometric mean) for the low-frequency (LF)-filtered complex tones. A sigmoid function was fitted to the HF data (upper black curve). Vertical dashed lines represent the F_0 transition point yielding the F_0DL halfway between the maximum and the minimum thresholds. The lower black curve depicts the mean of the LF data. Error bars represent the standard error of the mean.

frequency selectivity (Bernstein and Oxenham, 2006b). $F_{0,tr}$ occurred at similar F_0 s for musicians and non-musicians ($F_{0,tr, musicians} = 193$ Hz; $F_{0,tr, non-musicians} = 187$ Hz), suggesting that the two groups of listeners had similar auditory filter bandwidths. A one-way unbalanced ANOVA performed on the individual transition points for musicians and non-musicians revealed no significant difference in the mean between the two groups (mean \pm standard deviation 174 ± 45 Hz for musicians and 192 ± 30 Hz for non-musicians; $F(1,13) = 0.74$, $P = 0.405$). Overall, the findings of experiment 1 suggest that musical training enhances pitch discrimination of resolved and unresolved complex tones to the same extent. However, musicians did not show enhanced peripheral frequency selectivity (as estimated from the $F_{0,tr}$) as compared to non-musicians.

4.3.2 Effects of Harmonic Resolvability and Task Difficulty

Figure 4.3 depicts the mean pitch discrimination thresholds obtained in experiment 2 for 10 musicians (left panel) and 10 non-musicians (right panel). Pitch discrimination thresholds for the LF-filtered complex tones (open symbols connected via linear interpolation) were measured at the 75 % point on the psychometric function, and the obtained mean thresholds (1 % for musicians and 2 % for non-musicians) were similar to the thresholds obtained in experiment 1. Pitch discrimination thresholds for the HF-filtered complex tones (filled symbols) were measured at three different points on the psychometric function (diamonds 60 %; circles 75 %; triangles 90 %). The effect of increasing the tracked performance level from 60 to 90 % of correct responses increased the thresholds, on average, by a factor of 4.9 and 6.3 for musicians and non-musicians, respectively. Similar to the results obtained in experiment 1, thresholds for the non-musicians were, on average, larger than thresholds for musicians by a factor of 1.64. A mixed model with group, F_0 , and task difficulty as main effects and listeners as random factor nested in group was fit to the set of data and revealed a significant effect of the main factors (group $F(1,219) = 5.5$, $P = 0.031$; F_0 $F(2,219) = 85.06$, $P < 0.0001$; task difficulty $F(2,219) = 197.43$, $P < 0.0001$). The individual thresholds obtained in experiment 2 were used in experiment 3 to adjust for the difficulty level across listeners.

4.3.3 Pupil Dilations During Pitch Discrimination

In experiment 3, pupil dilations were recorded during a pitch discrimination task, where the difference in F_0 between reference and deviant was set at the individual thresholds obtained in experiment 2. This allowed for matching the difficulty level across listeners (60 % high task difficulty, 75 % medium task difficulty, 90 % low task difficulty). Figure 4.4 depicts the mean pupil dilation relative to baseline as a function of time (time zero refers to the beginning of stimulation with complex tones). For all conditions, the pupil dilated during stimulation with complex tones until it reached maximum dilation, on average at 1.78 s after stimulus onset for musicians and 1.87 s for non-musicians. After the maximum dilation point, pupil

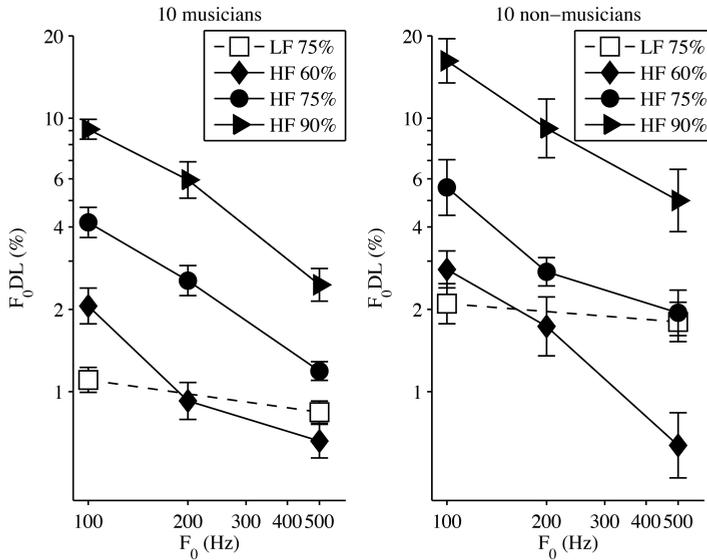


Figure 4.3: Mean pitch discrimination thresholds (F_0 DLs) as a function of F_0 , for 10 musicians (left panel) and 10 non-musicians (right panel). The filled symbols depict the thresholds (geometric mean) for the high-frequency (HF)-filtered complex tones (diamonds threshold at the 60 % point on the psychometric function; circles threshold at the 75 % point; triangles threshold at the 90 % point). The open squares depict the thresholds (geometric mean) for the low-frequency (LF)-filtered complex tones (threshold at the 75 % point on the psychometric function). All lines depict the linear interpolants between two consecutive thresholds. Error bars represent the standard error of the mean.

size decreased with longer decay times for non-musicians than for musicians until reaching the zero baseline value, on average at 3.2 s for non-musicians and at 2.8 s for musicians. As the largest effect of task difficulty occurred after the maximum dilation point, the time-averaged pupil size was calculated from the first occurring maximum dilation point (at 1.72 s) until 4.5 s after stimulus onset. The normalized mean values are presented in Figure 4.5, where the black, grey, and white bars depict the difficult, medium-difficult, and easy task condition, respectively. Results are presented for 10 musicians (left panels) and 10 non-musicians (right panels), at the three resolvability levels (top panels $F_0 = 100$ Hz, unresolved complex tones; middle panels $F_0 = 200$ Hz, mid-resolved tones; bottom panels $F_0 = 500$ Hz, resolved tones). Musicians had significantly smaller pupil dilations than non-musicians across

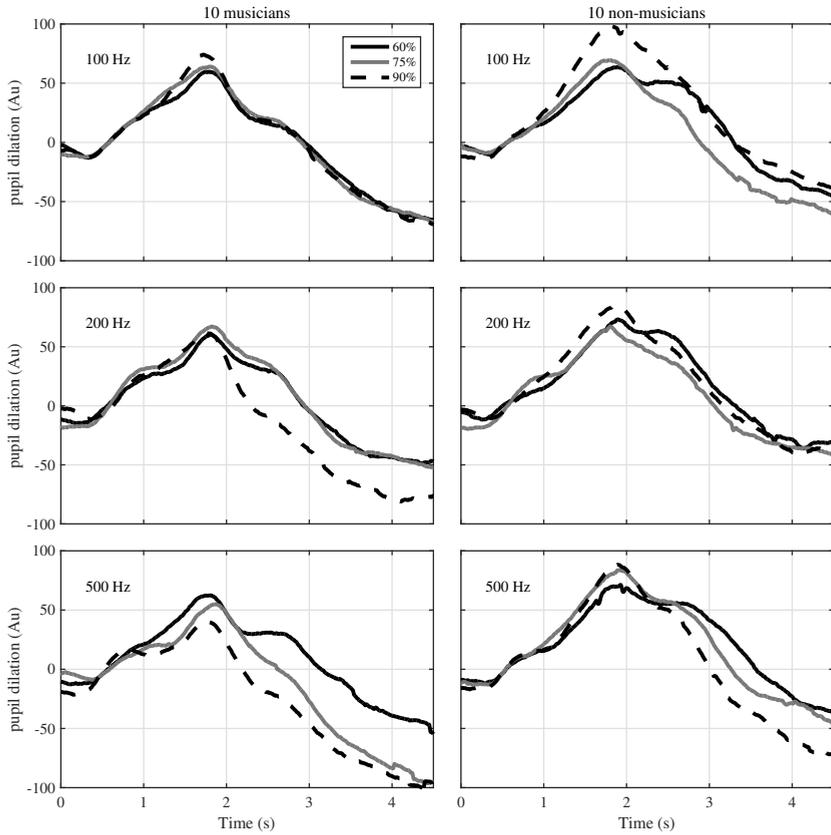


Figure 4.4: Mean pupil dilation [Au] for 10 musicians (left panels) and 10 non-musicians (right panels). The solid black, gray, and dashed curves represent pupil dilations at three task difficulty levels (60 % difficult task; 75 % medium difficulty; 90 % easy task). The top, middle, and bottom panels show pupil dilations for unresolved complex tones ($F_0 = 100$ Hz), mid-resolved tones ($F_0 = 200$ Hz), and resolved tones ($F_0 = 500$ Hz), respectively.

conditions (one-tailed unpaired t test $P = 0.031$), suggesting a lower processing effort for the same difficulty level. Ad hoc unpaired one-tailed t tests revealed that pupil dilations for musicians were smaller than dilations for non-musicians when the tones were resolved ($F_0 = 500$ Hz, bottom panels in Fig. 4.5) and the task was either medium-difficult ($P = 0.018$ with Bonferroni correction, asterisks above the grey bars) or easy ($P = 0.057$ with Bonferroni correction) and when the tones were mid-resolved ($F_0 = 200$ Hz, middle panels in Fig. 4.5) and the task was easy ($P = 0.003$ with Bonferroni correction, asterisks above white bars).

A mixed model with group, F_0 , and task difficulty as main effects and listeners as random factor nested in group was fit to the set of data and revealed a significant effect of task difficulty ($F(2,179) = 4.27$; $P = 0.016$) on pupil dilation. Ad hoc paired one-tailed t test revealed that there was a trend for pupil size to increase from the easy-task condition (white bar) to the difficult-task condition (black bar) for resolved complex tones ($F_0 = 500$ Hz, $P = 0.058$ with Bonferroni correction, bottom left panel in Fig. 4.5) and for the mid-resolved tones ($F_0 = 200$ Hz, $P = 0.024$ with Bonferroni correction, asterisk in the middle left panel in Fig. 4.5). When the complex tones were unresolved ($F_0 = 100$ Hz), pupil size was largely independent of the difficulty level. Although the analysis did not reveal a significant general effect of F_0 ($F(2,179) = 0.38$; $P = 0.687$), the interaction factor of F_0 and task difficulty was significant ($F(4,179) = 2.66$; $P = 0.035$). For the non-musicians, neither task difficulty nor resolvability had a significant effect on pupil dilation (two-factor ANOVA; difficulty $F(2, 89) = 0.87$, $P = 0.437$; resolvability $F(2,89) = 0.12$, $P = 0.890$), although a similar effect of task difficulty as for musicians occurred for the resolved stimuli ($F_0 = 500$ Hz). The two LF control conditions (at F_0 s of 100 and 500 Hz, listed in Figure 4.1) showed similar pupil dilation as the HF condition that matched in resolvability and task difficulty (one-way ANOVA with F_0 as main effect, $F(2,29) = 1.44$, $P_{\text{musicians}} = 0.254$; $F(2,29) = 1.47$, $P_{\text{non-musicians}} = 0.247$). Figure 4.6 depicts the correlation between the mean time-averaged pupil dilation and the percentage of correct responses, for musicians (filled squares) and non-musicians (open circles). The linear fit (dashed line in Fig. 4.6) to the musicians' mean data revealed a significant correlation between performance and pupil size ($P = 0.044$), whereby a decrease in performance was reflected in larger pupil dilations (i.e., larger effort). A

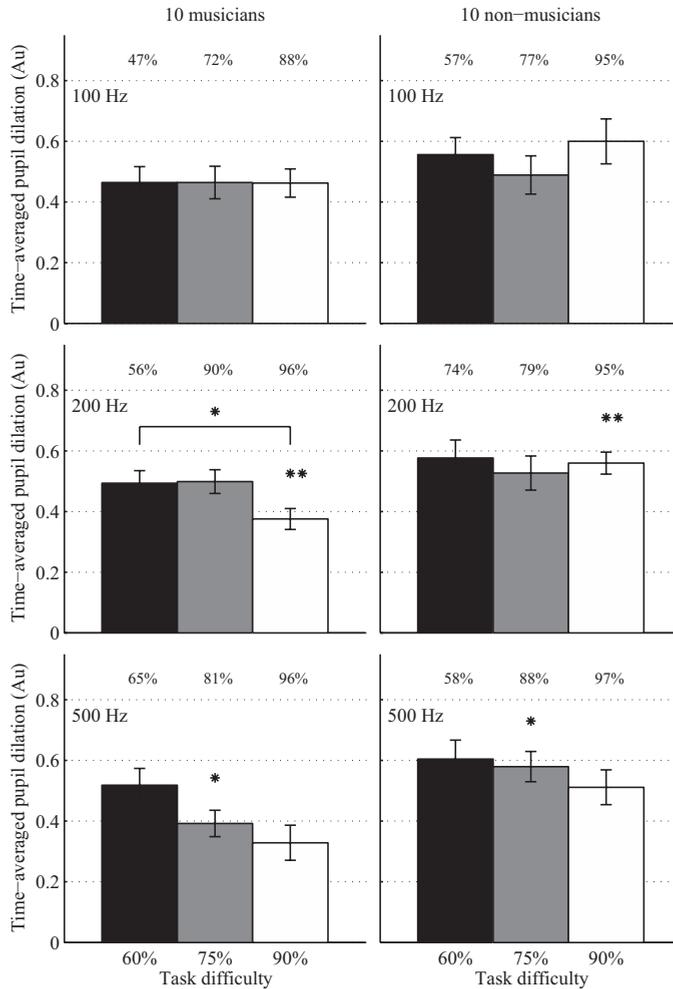


Figure 4.5: Mean normalized time-averaged pupil dilation (from maximum dilation until 4.5 s after stimulus onset), for 10 musicians (left panels) and 10 non-musicians (right panels). Normalization was done by subtracting the minimum pupil dilation (across all data) from the individual data and by dividing by the maximum range. The black, grey, and white bars represent pupil dilations at three task difficulty levels (60 % difficult task; 75 % medium difficulty; 90 % easy task). The percentages reported on the upper portion of each panel represent the average of correct responses across listeners in each condition. The top, middle, and bottom panels show pupil dilations for unresolved complex tones ($F_0 = 100$ Hz), mid-resolved tones ($F_0 = 200$ Hz), and resolved tones ($F_0 = 500$ Hz), respectively. Asterisks depict the conditions for which one-tailed t tests reported significance (* $P \leq 0.05$; ** $P \leq 0.01$, with Bonferroni correction). Error bars represent the standard error of the mean.

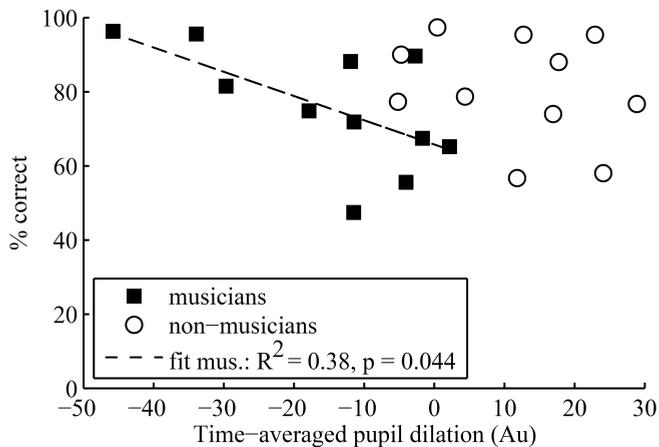


Figure 4.6: Correlation between the mean time-averaged (from maximum dilation until 4.5 s after stimulus onset) and baseline-corrected pupil dilation (in arbitrary units, Au) and the percentage of correct responses, for 10 musicians (filled squares) and 10 non-musicians (open circles) in all the 11 tested conditions. A linear model was fit to the mean data of musicians (dashed line).

decrease in performance below 65 % did not lead to a further increase in pupil size, which may indicate a decrease in processing effort following a too demanding task (i.e., cognitive processing overload). No trend between performance and pupil dilations was observed in non-musicians. Figure 4.7 depicts the mean reaction times for button press for musicians and non-musicians in all 11 tested conditions. Listeners pressed the response button, on average, 1 s after stimulus offset. A mixed model with group, F_0 , and task difficulty as main effects and listeners as random factor nested in group was fit to the set of data and revealed no significant difference in reaction times across the two groups of listeners ($F(1, 208) = 0.0024$; $P = 0.961$), while both F_0 and task difficulty had a significant effect on the reaction times ($F_0 F(2, 208) = 8.32$, $P = 0.0003$; difficulty $F(2, 208) = 73.66$, $P < 0.0001$). This finding confirmed that the slower decay time of pupil dilations in non-musicians versus musicians was not an effect of longer reaction times in non-musicians but rather indicated a larger processing effort in performing the task with increasing task difficulty and decreasing harmonic resolvability.

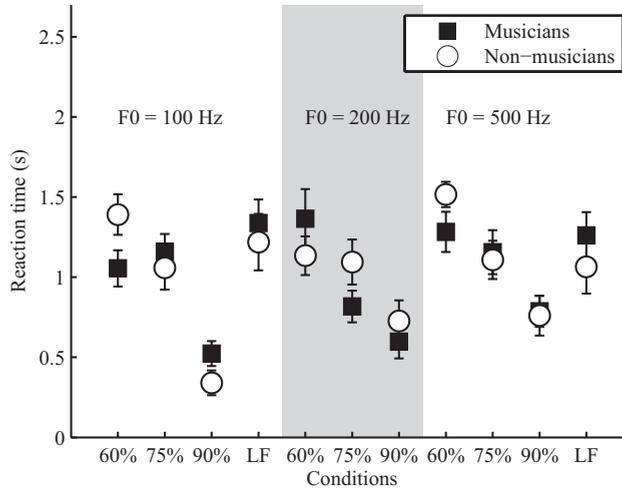


Figure 4.7: Mean reaction time (time in seconds from stimulus offset to button press) for musicians (filled squares) and non-musicians (open circles), for all 11 tested conditions. Error bars represent the standard error of the mean.

4.4 Discussion

In a first behavioral experiment, pitch discrimination thresholds for resolved and unresolved complex tones were measured in musicians and non-musicians. The findings of experiment 1 (Fig. 4.2) revealed pitch discrimination thresholds similar to those reported by Bernstein and Oxenham (2006b), whereby the thresholds for the HF-filtered complex tones decreased with increasing harmonic resolvability. Moreover, the current findings suggest that musical training improved pitch discrimination of resolved and unresolved complex tones to a similar extent. The difference in performance between the two groups of listeners was, on average, of about a factor of 1.72. This value was similar to the enhancement reported by previous studies in pitch discrimination thresholds of pure tones in musically trained listeners (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001). Since the exact extent of the enhancement was shown to depend on the selection criterion of the musically trained listeners and on the amount of training (Micheyl et al., 2006), the

current study did not focus on quantifying the difference in performance between the two groups but rather on comparing the enhancement between resolved and unresolved complex tones. The rationale behind this was that if musicians had a higher peripheral frequency selectivity, as suggested by Bidelman et al., 2014; Bidelman et al., 2016, pitch discrimination thresholds would show a larger enhancement in performance for resolved versus unresolved complexes and, additionally, the transition point ($F_{0, \text{tr}}$) at which components would become resolved would occur at smaller F_0 s in musicians. As the current findings showed not only a similar enhancement for resolved and unresolved complexes but also a similar $F_{0, \text{tr}}$ for the two groups of listeners, the results of experiment 1 suggest similar peripheral frequency selectivity in musicians versus non-musicians. This finding does not rule out a possible finer representation of F_0 at higher stages of the auditory system in musicians. In fact, while $F_{0, \text{tr}}$ is considered to reflect a peripheral limitation of the auditory filters to resolve the individual harmonics (Bernstein and Oxenham, 2006b), a finer F_0 representation in musicians might still occur at more central stages of the auditory system (e.g., at stages after F_0 extraction) and affect pitch discrimination thresholds of both resolved and unresolved complexes, without necessarily affecting the transition point. This interpretation of the results would additionally be supported in the context of pitch perception involving different mechanisms for resolved and unresolved harmonics. In fact, if the pitch discrimination enhancement in musicians occurred at stages of the auditory system preceding F_0 extraction, different enhancements would be expected to occur for resolved and unresolved harmonics. Thus, the almost identical sizes of the differences (expressed as ratios) in thresholds between musicians and non-musicians for resolved and unresolved harmonics suggests a training-dependent enhancement in musicians that is independent of the pitch extraction mechanism and likely to occur centrally in the auditory system (e.g., a finer cortical F_0 representation). In experiment 3, processing effort was investigated via pupil dilation in musicians and non-musicians. The pupil size was recorded during a pitch discrimination task for conditions at three levels of resolvability (unresolved, mid-resolved, and resolved complex tones) and task difficulty (high, medium, and low difficulty). The results (Fig. 4.5) revealed that pupil dilations in musicians were lower than in

non-musicians in all conditions. As an increase of pupil size has in previous studies been shown to reflect an increase in processing effort (e.g., Janisse, 1977; Beatty, 1982), lower dilations in musicians suggest a lower effort in performing the task, although the difficulty level was matched across the two groups of listeners. Thus, at similar (i.e., individually adjusted) processing demands imposed by the pitch discrimination task, it was still less demanding to extract pitch-related features for musically trained listeners. Interestingly, dilations were significantly lower in musicians when the complex tones were resolved and the task difficulty was either low or medium (asterisks above grey and white bars in Fig. 4.5). A mixed model with three factors (resolvability, difficulty, group) confirmed a significant interaction of both group and difficulty ($F(2,219) = 3.26; P = 0.05$) and of resolvability and difficulty ($F(2,219) = 2.61; P = 0.043$). The fact that dilations were significantly lower in musicians versus non-musicians for resolved but not for unresolved complexes may indicate either an increased ability to extract the pitch of resolved stimuli following musical training or an increased sensitivity along the auditory pathway to resolved stimuli in musicians (e.g., a finer cortical representation). Moreover, pupil dilations were significantly correlated with behavioral performance in musicians (Fig. 4.6), whereby a decrease in performance from 96 to 65 % was reflected in a progressive increase of pupil dilations. When the performance was lower than 65 %, a drop in pupil dilations was observed in musicians, which may suggest a cognitive processing overload. Previous studies recording pupil dilations during performance of cognitive tasks also reported a decrease in pupillary responses when the task processing demands exceeded the listener's processing resources (Granholm et al., 1996; Zekveld and Kramer, 2014). For non-musicians, neither task difficulty nor resolvability had a significant effect on pupil dilation. Additionally, pupil dilation for the condition with lowest processing demand (i.e., condition of low task difficulty and high resolvability) did not differ (paired t test: $P = 0.382$) from the dilation for the condition with highest processing demand in non-musicians. This might indicate a ceiling effect in non-musicians, whereby already the condition with lowest processing demand approached the available cognitive resources allocated for pitch discrimination not allowing for a further increase in pupil dilations when either increasing the task difficulty or decreasing the resolvability of

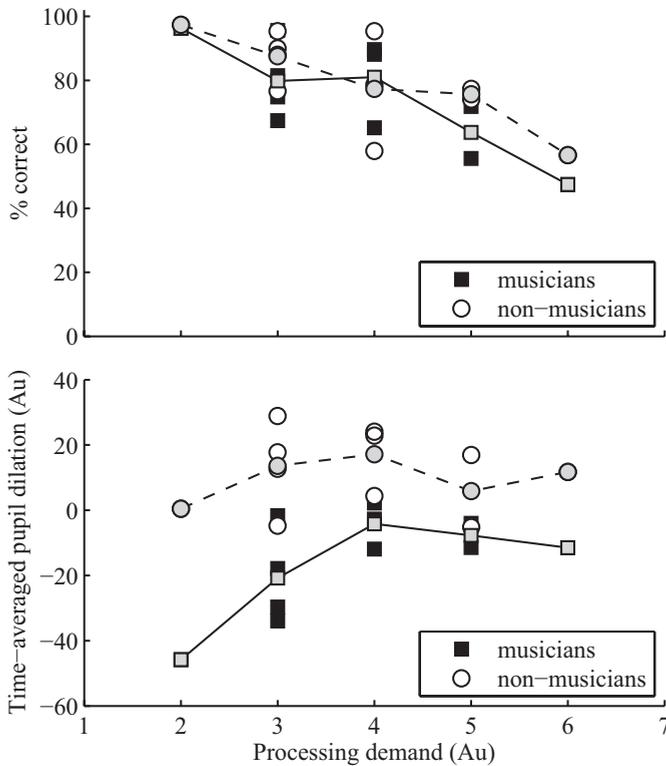


Figure 4.8: Behavioral performance (% correct deviant identification, top panel) and time-averaged pupil dilation (bottom panel) as a function of the processing demand of the 11 presented conditions. Filled black squares and open circles depict all individual conditions for musicians and non-musicians, respectively. Grey squares and grey circles depict the mean values for conditions of equal processing demand for musicians and non-musicians, respectively. Processing demand is calculated as the sum of arbitrary weights assigned for both task difficulty (1 easy task; 2 medium-difficult task; 3 difficult task) and harmonic resolvability (1 resolved tones; 2 medium resolved tones; 3 unresolved tones).

the stimuli.

As the three-factor ANOVA revealed an interaction between task difficulty and resolvability on pupil responses, the obtained dilations were additionally related to

the overall processing demand imposed to the listener by the combination of these two factors in each listening condition (Johnsrude and Rodd, 2015). Processing demand was calculated as the sum of arbitrary weights on a scale from 1 to 3 (1 low processing demand; 2 medium processing demand; 3 high processing demand), assigned for both task difficulty (1 low task difficulty; 2 medium task difficulty; 3 high task difficulty) and harmonic resolvability (1 resolved tones; 2 medium resolved tones; 3 unresolved tones). Thus, a condition with resolved complex tones and an easy task would impose to the listener the lowest processing demand (i.e., a total weight of 2), while a condition with unresolved complex tones and a difficult task would impose the highest processing demand (i.e., a total weight of 6). Figure 4.8 depicts behavioral performance (top panel) and time-averaged pupil dilation (bottom panel) as a function of the processing demand of the 11 presented conditions, for musicians (black squares individual conditions; grey squares mean of conditions with equal processing demand) and non-musicians (open circles individual conditions; grey circles mean of conditions with equal processing demand). The solid and dashed lines (top panel) depict the linear interpolant to the mean data for musicians and non-musicians, respectively. A linear fit to the data indicated a significant negative correlation between behavioral performance and processing demand imposed by each condition (musicians $R^2 = 0.93$, $P = 0.008$; non-musicians $R^2 = 0.94$, $P = 0.006$). Additionally, a mixed model with group, F_0 , and task difficulty as main effects and listeners as random factor nested in group was fit to the set of data and revealed no significant difference in behavioral performance across the two groups of listeners ($F(1, 208) = 1.63$; $P = 0.22$), in agreement with the experimental design that was built to match the task difficulty across listeners. Although musicians and non-musicians performed similarly in the presented conditions, the amount of processing effort to compensate for processing demand differed markedly (bottom panel in Fig. 4.8). While for the musicians pupil dilation increased with increasing processing demand until reaching a plateau (solid line in bottom panel of Fig. 4.8), consistent with Johnsrude and Rodd, 2015, pupillary responses approached a plateau value already for conditions imposing the lowest processing demand (dashed line) for the non-musicians. This finding is in agreement with previous studies investigating pupillary responses

during different types of cognitive tasks, where it was found that pupil dilation increases with increasing task processing demands until reaching resource limits (Pook, 1973; Granholm et al., 1996; Johnsrude and Rodd, 2015). This plateau value is maintained as long as the listener is able to allocate maximal processing resources, after which pupil dilation decreases as a result of a resource overload condition (Pook, 1973; Granholm et al., 1996).

4.5 Conclusion

Overall, the findings of the current study revealed a similar enhancement in pitch discrimination of resolved and unresolved complex tones in the musically trained listeners compared to the non-musicians. This enhancement is unlikely to be related to higher peripheral frequency selectivity in the musicians, since the improved performance was not specific to only resolved complex tones and, additionally, the transition point for resolvability occurred at similar F_0 s in the musicians and non-musicians. An overall shift of the pitch discrimination thresholds might thus be related to a higher general ability to extract pitchrelated features following musical training and/or to a finer F_0 representation at more central stages of the auditory system. Pupillometry responses indicated a lower processing effort in the musicians versus the non-musicians, although the processing demand imposed by the pitch discrimination task was individually adjusted according to the behavioral thresholds. Thus, although the task difficulty was adjusted to compensate for the higher pitch discrimination thresholds in the non-musicians, the non-musically trained listeners still allocated higher cognitive resources than did the musicians to perform the task at the same performance level (% correct). This finding might suggest an enhanced pitch representation along the auditory system in musicians and possibly a finer F_0 representation at central stages of the auditory system. Future work may clarify this hypothesis by investigating pitch representations in the auditory cortex in musicians versus non-musicians via functional magnetic resonance imaging.

Acknowledgments

This work was supported by the Technical University of Denmark. The authors would like to thank Robert Zatorre for the interesting discussions on the behavioral results and Wiebke Lamping for her help in recruiting the listeners and performing the behavioral and pupillometry measurements (experiments 2 and 3).

5

Cortical correlates of complex-tone pitch discrimination in musicians and non-musicians^a

Abstract

Musicians have been shown to have an enhanced pitch-discrimination ability compared to non-musicians for complex tones with either resolved or unresolved harmonics. It is unclear whether this perceptual enhancement can be ascribed to an enhanced neural representation of pitch at central stages of the auditory system. The aim of this study was to clarify whether (i) cortical responses increase with harmonic resolvability, as suggested in previous studies, and whether musicians show (ii) differential neural activation in response to complex tones as compared to non-musicians and/or (iii) finer fundamental frequency (F_0) representation in the auditory cortex. Assuming that the right auditory cortex is specialized in processing fine spectral changes, we hypothesized that an enhanced F_0 representation in musicians would be associated with a stronger right-lateralized response to complex tones compared to non-musicians. Fundamental frequency (F_0) discrimination thresholds (F_0 DLs) were first estimated behaviorally in musicians and non-musicians for harmonic complex tones with F_0 s of 100 and 500 Hz, filtered in either a low or a high frequency region to

^a This chapter is based on Bianchi, F., Hjortkjær, J., Santurette, S., Siebner, H., Zatorre, R., Dau, T., (in preparation for J. Neurosci.).

vary the resolvability of audible harmonics. In a second experiment, a sparse-sampling event-related functional magnetic resonance imaging (fMRI) paradigm was used to measure neural activation while the listeners performed the same pitch-discrimination task for conditions of varying resolvability. The task difficulty was individually adjusted according to the previously obtained F_0 discrimination thresholds. The behavioral results showed that the pitch-discrimination thresholds of musicians were lower than the thresholds of non-musicians for all conditions, with a stronger benefit for the resolved conditions. A group analysis on the fMRI results revealed no differential neural activation for resolved *vs.* unresolved conditions, suggesting that cortical responses did not increase with increasing stimulus resolvability when adjusting for the task difficulty across conditions and participants. Additionally, the group analysis revealed larger neural activation in the musicians relative to the non-musicians in the right Heschl's gyrus, right insula, right middle and superior frontal gyri and inferior colliculus. Finally, neural responses in the right auditory cortex were predictive of the individual pitch-discrimination thresholds only in the musically trained listeners, consistent with a higher specialization of the right auditory cortex in processing fine spectral changes relative to the left auditory cortex and with a training-dependent plasticity in musicians. Overall, these findings suggest an increasing activation of the right-lateralized pitch-sensitive cortical areas with increasing musical abilities.

5.1 Introduction

Natural sounds, like speech and music, typically contain complex harmonic structures that elicit a pitch corresponding to the fundamental frequency (F_0) of the sound stimulus (e.g., Licklider, 1951; Schouten et al., 1962; De Cheveigné, 2005). Musicians are more trained than non-musicians to listen for and retrieve the pitch

of such complex stimuli. Possibly as a result of training, musicians have been shown to be more sensitive than non-musicians to discriminate fine spectral changes between complex tones. In fact, trained musicians have been found to have two to six times lower F_0 -discrimination thresholds than non-musicians (Spiegel and Watson, 1984; Micheyl et al., 2006; Allen and Oxenham, 2014; Bianchi et al., 2016b). This benefit was shown to depend on the overall duration of musical training, on the age when musical education started, as well as on the family of played instruments, with a smaller benefit for the musicians playing keyboard instruments as compared to strings and winds (Spiegel and Watson, 1984; Micheyl et al., 2006). Additionally, the musicians' benefit for complex-tone pitch discrimination was shown to partially extend to pure tones (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001), although the benefit for pure-tone pitch discrimination was smaller than for complex tones (Micheyl et al., 2006).

In a previous study, the musicians' advantage for complex-tone pitch discrimination was investigated as a function of harmonic resolvability to clarify whether resolved and unresolved harmonics contributed similarly to this benefit (Bianchi et al., 2016b). It was found that the musicians' benefit in pitch discrimination was similar for complex tones containing either resolved or unresolved harmonics. This finding suggested an enhanced F_0 representation along the auditory system in musicians independent of harmonic resolvability, and possibly occurring at stages after F_0 extraction. The motivation for the present study was to clarify whether this perceptual enhancement of musicians for complex-tone pitch discrimination containing either resolved or unresolved harmonics can be reflected by an enhanced neural representation of pitch at central stages of the auditory system.

Cortical responses to resolved and unresolved complex tones have been investigated in previous neuroimaging studies (Penagos et al., 2004; Hall and Plack, 2009; Garcia et al., 2010; Barker et al., 2011; Norman-Haignere et al., 2013). Although there is no general consensus on the exact anatomical location of a pitch center, it has been suggested that cortical pitch-sensitive regions are located in anterior regions of the auditory cortex and respond more strongly to complex tones with resolved harmonics as compared to complex tones containing only unresolved harmonics (Penagos et al., 2004; Norman-Haignere et al., 2013). This finding is

consistent with neurophysiological studies reporting that pitch-sensitive neurons in the anterolateral border of primary auditory cortex respond in proportion to pitch salience (Bendor and Wang, 2005; Fishman et al., 2013). The effect of musical training on the neural representation of pitch has also been widely investigated (e.g., Pantev et al., 1998; Schneider et al., 2002; Musacchia et al., 2007; Pantev and Herholz, 2011; Herholz et al., 2012), whereby functional and anatomical changes were reported to occur along the auditory pathway of musically trained listeners. However, it is still unknown how musical training affects the cortical and subcortical pitch representations of complex tones containing either resolved or unresolved harmonics.

The aim of this study was to clarify whether (i) cortical responses increase with harmonic resolvability similarly in musicians and non-musicians, and whether (ii) musicians show an increased cortical neural activation in response to complex tones and/or a finer F_0 representation along the auditory pathway. As previous studies have suggested that the right auditory cortex is more specialized than the left auditory cortex in processing fine spectral changes (e.g., Zatorre, 1988; Johnsrude et al., 2000; Zatorre and Belin, 2001; Zatorre et al., 2002; Hyde et al., 2008), it was hypothesized here that an enhanced cortical F_0 representation in musicians would be associated with a stronger right-lateralized response to complex tones compared to non-musicians.

Two experiments were carried out. In a first behavioral experiment, pitch-discrimination thresholds were measured for complex tones containing either resolved or unresolved harmonics to estimate the musicians' benefit in pitch-discrimination performance relative to the non-musicians. In a second experiment, functional magnetic resonance imaging (fMRI) was performed on the same listeners during a similar pitch-discrimination task, where the task difficulty was adjusted according to the individual thresholds from the first experiment and, thus, matched across participants. This allowed for disentangling the functional changes in the F_0 representation between musicians and non-musicians from changes in the difficulty to perform the task between the two groups of listeners.

5.2 Method

5.2.1 Listeners

Thirty-one listeners, aged from 22 to 30 years old, participated in this study. Sixteen participants (ten females) were musically trained listeners that had at least eight years of formal musical education. Fifteen participants (seven females) had no formal musical education and had never played a musical instrument. The two groups of listeners were matched in age (mean and median of musicians: 26 years old; mean and median of non-musicians: 25 years old). All participants were right handed according to the Edinburgh Handedness Inventory, except one musician who reported to be ambidextrous and was therefore excluded from Experiment II. All listeners had hearing thresholds of less than 20 dB hearing level (HL) at all audiometric frequencies between 125 Hz and 8 kHz. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark and were conducted in accordance with the Declaration of Helsinki.

5.2.2 Experiment I: Behavioral pitch-discrimination of complex tones

The ability to discriminate the pitch of resolved and unresolved complex tones was assessed via difference limens for fundamental frequency (F_0 DLs) as a function of F_0 . The procedure and the stimuli used here were similar to the ones used in Bianchi et al. (2016).

Procedure

A three-alternative forced choice (3-AFC) paradigm was used in combination with a weighted up-down method (Kaernbach, 1991) to track different points on the psychometric function (60%, 75% and 90%). For each trial, two intervals contained a reference complex tone with a fixed fundamental frequency ($F_{0,ref}$: 100 or 500 Hz) and one interval contained a deviant complex tone with a larger F_0 ($F_{0,ct}$). The initial difference in F_0 between reference and deviant, ΔF_0 ($F_{0,dev} - F_{0,ref}$) / $F_{0,ref}$,

was set to 20% and was then logarithmically decreased by a varying step size every second reversal. For each run, $F_{0,\text{ref}}$ was roved from trial to trial from a $\pm 5\%$ uniform distribution around the nominal value. A random level perturbation of ± 2.5 dB was applied to each interval, to prevent the listener from using loudness as a cue. The listener's task was to select the interval containing the deviant tone with a higher pitch than the two references. The threshold for each run was obtained as the geometric mean of the last six reversals. Before the actual test, the listeners performed three repetitions as training. The final value of $F_0\text{DL}$ was calculated from the mean of three repetitions.

Stimuli

All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 300-ms complex tones with harmonic components added in sine phase and embedded in broadband threshold equalizing noise (TEN, Moore et al., 2000). The sound pressure level (SPL) of the TEN was set to 45 dB per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) to mask the combination tones. The level of each harmonic component was fixed at 50 dB SPL. Figure 5.1(a) depicts the conditions used in this study. Conditions of varying resolvability were achieved by bandpass filtering the complexes in a high-frequency region (HF, 1500-3500 Hz, bottom panels of Fig. 5.1(a)), with 50 dB/octave slopes, and by using an F_0 of either 100 Hz (unresolved condition) or 500 Hz (resolved condition). Two control conditions with complexes filtered in a low-frequency region (LF, 300-1500 Hz, top panels of Fig. 5.1(a)) and F_0 s of either 100 or 500 Hz (resolved conditions) were used to control for changes in fundamental frequency (Penagos et al., 2004). For the HF-filtered complexes, two different points on the psychometric function were estimated at 60% and 90% correct performance. The 60% point corresponded to a difficult pitch-discrimination task, and the 90% point corresponded to an easy task. For the LF-filtered complexes, only the 75% point on the psychometric function was estimated, resulting in a task of medium difficulty. Thus, six conditions were tested in total (summarized in Fig. 5.1(b)): 100-Hz HF (60% and 90%; unresolved conditions), 500-Hz HF (60% and 90%; resolved conditions), 100-Hz LF (75%;

resolved condition) and 500-Hz LF (75%; resolved condition). The stimuli were presented diotically through equalized headphones (Sennheiser HD 650).

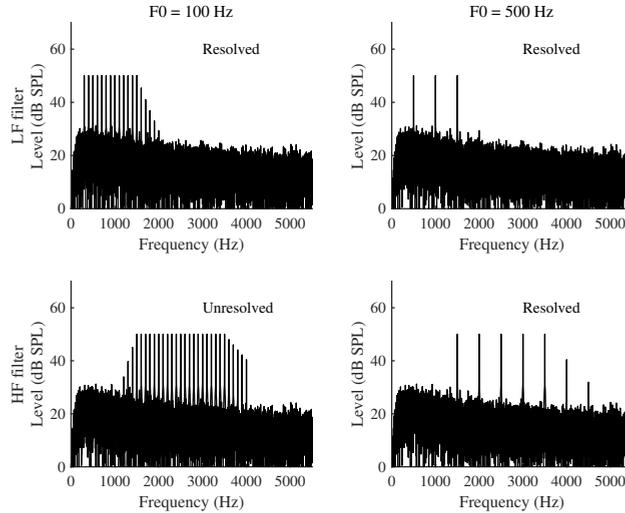
5.2.3 Experiment II: Cortical responses during a pitch-discrimination task

Imaging protocol

Functional imaging was performed on a 3 tesla scanner (Philips Achieva with 32-channel head coil) using a T2*-weighted echo-planar imaging sequence (TR = 10 sec, TE = 30 msec; flip angle, 90°). Thirty-eight slices (slice thickness of 3 mm; isotropic voxel size of 3x3x3 mm³) oriented parallel to the lateral sulcus and covering the entire brain were imaged. The acquisition time of one volume was of 2.5 s (see Fig. 5.2). A sparse imaging sequence (Hall et al., 1999) was used, where the sound stimuli were presented in the silent period between two volume acquisitions. T1-weighted anatomical images (1x1x1 mm³) were also acquired for each participant.

Procedure

For each trial, two stimuli were reference complex tones with a fixed fundamental frequency (100 or 500 Hz) and one stimulus was a deviant complex tone with a larger F_0 (denoted by the asterisk in Fig. 5.2). The deviant position was randomized across trials and runs. The first tone was presented 2 to 3 seconds after the acquisition of the previous volume. This time jitter in the onset of the signal was introduced to account for the inter-subject variability of the blood-oxygen-level dependent (BOLD) hemodynamic response (Aguirre et al., 1998). Participants had to identify the deviant tone by pressing either the first, second or third button on a mouse, according to the deviant's position (i.e., first, second or third tone). The participants were instructed to press the response button during the following volume acquisition (see Fig. 5.2). This allowed for disentangling the BOLD signal relative to the pitch-discrimination task and to the button press. In a control experiment run in a soundproof booth, it was checked that pressing the response button 2 to 3



(a)

		F_0	
		100 Hz	500 Hz
task difficulty	90%	Unresolved Easy	Resolved Easy
	75%	Resolved Medium difficult	Resolved Medium difficult
	60%	Unresolved Difficult	Resolved Difficult

(b)

Figure 5.1: a) Stimuli used in Experiment I and II: complex tones with an F_0 of either 100 Hz or 500 Hz, filtered in either a low-frequency region (LF: 0.3-1.5 kHz) or high-frequency region (HF: 1.5-3.5 kHz) and embedded in TEN. b) Summary of the six pitch conditions tested in Experiment I and II. Complex tones filtered in a LF region are depicted in blue and complex tones filtered in a HF region are depicted in red. Task difficulty was varied by adjusting ΔF_0 according to the individual pitch-discrimination thresholds at the 60% (high difficulty), 75% (medium difficulty) and 90% (low difficulty) points on the psychometric function.

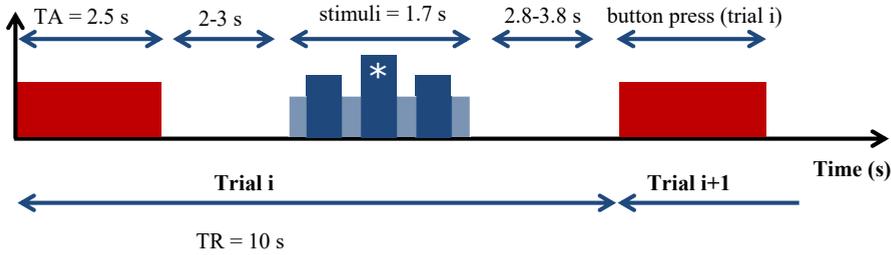


Figure 5.2: Schematic of the stimulus presentation for Experiment II. The red blocks depict the acquisition of one volume (acquisition time, TA of 2.5 s). The three complex tones (blue bars) were embedded in noise and were presented 2 to 3 seconds after the acquisition of the previous volume. The deviant position (asterisk) was randomized across trials. The participants were instructed to press the response button during the following volume acquisition.

seconds after the stimulus presentation and in the presence of scanner noise did not affect the participants' performance. The difficulty of the pitch-discrimination task was defined by the difference in F_0 between reference and deviant (i.e., ΔF_0), which was adjusted for each participant according to the individual thresholds measured in Experiment I (60%: high difficulty; 75%: medium difficulty; 90%: low difficulty). The seven stimulus conditions (six pitch conditions and one noise condition) were randomly presented six times in each run, for a total of 42 trials per run (about 7 min). In total, six runs were carried out in the same scanning session (about one hour).

Stimuli

The same conditions as in Experiment I were used for the fMRI paradigm. The six pitch conditions (four HF-filtered complex tones and two LF-filtered complex tones, see Fig. 5.1(b)) with a level of 50 dB SPL per harmonic and embedded in TEN at 45 dB SPL/ERB were randomly presented to the participants during the inter-scan interval. Additionally, a noise condition with broadband TEN (45 dB SPL/ERB) was used as a baseline condition. All conditions were 1.7 s long. The sound stimuli were presented diotically through equalized MRI-compatible insert earphones (Sensimetrics S14).

Data Analysis

Data analyses were performed with the statistical parametric mapping software (SPM8, Wellcome Trust Centre for Neuroimaging, London, UK). Data processing consisted of realignment, coregistration, spatial normalization to MNI standard space as implemented in SPM8, and smoothing with an 8 mm full-width at half-maximum isotropic Gaussian kernel. Data analysis was performed using a general linear model (GLM) approach. At the individual level (first level analysis), separate regressors were defined for each experimental condition (seven regressors) to model the onset of the sound stimulus. Six regressors with the realignment parameters were also used. Low frequency drifts in the BOLD signal were removed by a high-pass filter with a cut-off period of 128 s.

For the group analysis (second level analysis), a full-factorial ANOVA was fit to the set of data. The design matrix included three main factors: task difficulty (three levels: 60%, 75%, 90%), F_0 (two levels: 100 and 500 Hz) and group (musicians and non-musicians). Two covariates were included in the design matrix: the ΔF_0 at which each condition was tested (i.e., the difference in F_0 between reference and deviant measured in Experiment I) and the percent of correct deviant identification. The mean T1-weighted image was calculated across 25 listeners, since the T1-weighted images of the remaining 5 listeners was affected by eye-movements artifacts. Voxelwise FWE corrected p-values reported in this study were obtained from whole brain analysis.

A region of interest (ROI) analysis was carried out to identify the pitch-sensitive voxels within the right and left auditory cortices (Norman-Haignere et al., 2013). An anatomical mask, comprising primary and non-primary auditory cortices (Te1.0, Te1.1, Te1.2 and Te3, see Fig. 5.3(a)), was built using the Anatomy toolbox from SPM8. The 10% most activated voxels for all pitch conditions relative to the noise were selected within this anatomical mask for each participant. In an independent analysis, the mean response of these pitch-sensitive voxels was estimated for each pitch condition relative to the noise (four resolved conditions and two unresolved conditions).

In order to relate the functional activation in frontal regions with the listeners'

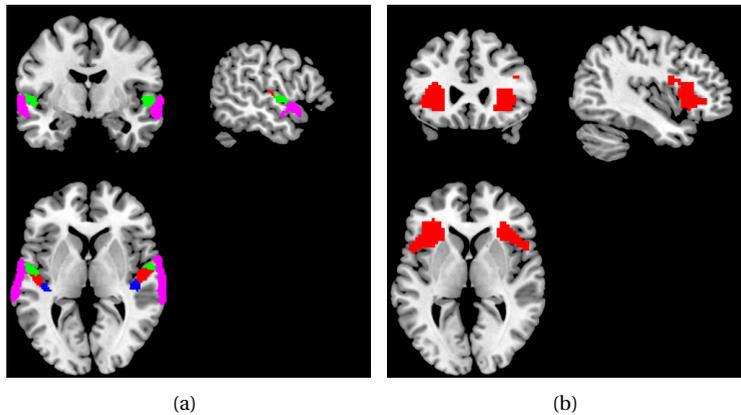


Figure 5.3: a) Anatomical mask comprising primary (Te1.0, Te1.1, Te1.2) and non-primary (Te3) auditory cortex. Posteromedial Heschl's gyrus is shown by the red (Te1.0) and blue (Te1.1) regions (the "core" areas, Hall and Barker, 2012), antero-lateral Heschl's gyrus is depicted by the green region (Te1.2). Non-primary auditory cortex is shown in pink (Te3, planum temporale). b) Functional frontal mask, comprising the insula and inferior frontal operculum, obtained from the contrast 60%>90% ($p < 0.001$ uncorrected).

performance (i.e., % correct deviant identification), a functional mask was used (Fig. 5.3(b)). This mask, comprising the insula and inferior frontal operculum, included all thresholded voxels ($p < 0.001$ uncorrected) obtained from the contrast 60%>90% (from the full-factorial ANOVA without the covariates). This frontal mask was applied to the individual activation maps obtained for each pitch condition relative to the noise. The 10% most activated voxels within the inclusive mask were selected for each participant and used to clarify the existence of a correlation with performance.

5.3 Results

5.3.1 Experiment I: Pitch-discrimination of complex tones

Figure 5.4(a) depicts the mean pitch-discrimination thresholds for the four conditions with resolved harmonics (left panel) and the two conditions with unresolved

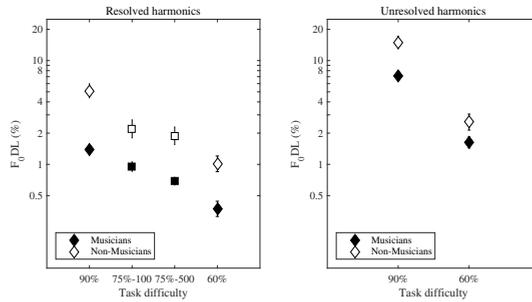
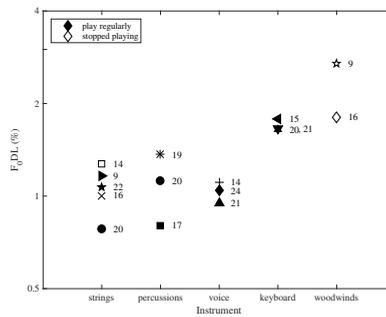
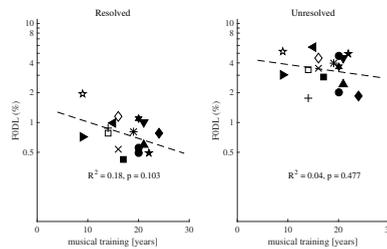
(a) Mean F_0DL s for musicians and non-musicians.(b) Individual F_0DL s for the musicians grouped according to their main instrument.(c) Individual F_0DL s for the musicians as a function of years of training.

Figure 5.4: a) Mean pitch-discrimination thresholds for the 16 musicians (filled symbols) and 15 non-musicians (open symbols), for the four resolved conditions tested at the 90%, 75% and 60% points on the psychometric functions (left panel) and the two unresolved conditions tested at the 90% and 60% points on the psychometric function (right panel). Error bars depict the standard error of the mean. b) Individual pitch-discrimination thresholds averaged across the six tested conditions for the 16 musicians, divided into different categories according to the family of their main instrument. The three open symbols depict the musicians who stopped playing regularly. The numbers next to the symbols indicate the overall years of musical training. c) Individual pitch-discrimination thresholds averaged across the four resolved conditions (left panel) and the two unresolved conditions (right panel) for the 16 musicians, as a function of the overall years of musical training. Same symbols as Fig. 5.4(c).

harmonics (right panels) for the musicians (filled symbols) and non-musicians (open symbols). Compared to non-musicians, the musically trained listeners had significantly lower thresholds in all conditions, indicating a more accurate pitch-discrimination performance for both resolved and unresolved complex tones. The performance for all listeners increased in the presence of resolved harmonics, consistent with a more salient pitch percept evoked by the resolved than the unresolved harmonics (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2006b). Additionally, as expected from estimating a lower point on the psychometric function, the thresholds of both musicians and non-musicians decreased from the 90% to the 60% condition. A mixed-model ANOVA with three fixed factors (group, resolvability and difficulty) and listeners as a random factor nested in group was performed on the results. A significant effect of the three main factors was found: group [$F(1, 185) = 21.76$; $p = 0.0001$], resolvability [$F(1, 185) = 364.75$; $p < 0.0001$] and task difficulty [$F(2, 185) = 268.35$; $p < 0.0001$], as well as a significant interaction between group and resolvability [$F(1, 185) = 17.16$; $p = 0.0001$] and group and difficulty [$F(2, 185) = 3.86$; $p = 0.0233$]. In fact, the musicians' performance for the resolved conditions was enhanced, on average, by a factor of 2.8 relative to the non-musicians' performance and only by a factor of 1.8 for the unresolved conditions. The effect of musical training was also greater for the easy-task conditions (90% point of the psychometric function) as compared to the difficult-task conditions (60%).

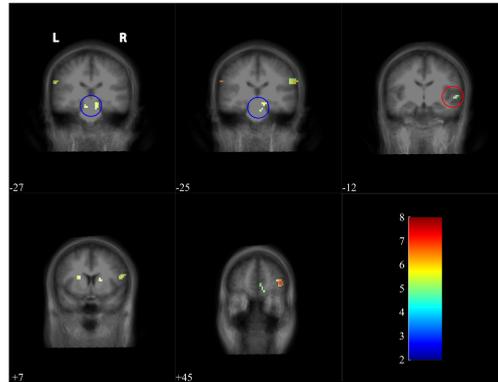
Figure 5.4(b) depicts the individual pitch-discrimination thresholds, averaged across all six conditions (resolved and unresolved conditions), for each musically trained participant. The musicians were assigned to a family of musical instruments according to the family of their main instrument. The worst performance was obtained by the musicians who played a keyboard instrument (Micheyl et al., 2006) or woodwinds, although both listeners playing woodwinds were also the ones who stopped playing regularly (open symbols). Musicians playing string instruments, classical percussions or singing reached the most accurate performance. Figure 5.4(c) depicts the individual pitch-discrimination thresholds for the 16 musicians, averaged across the four resolved conditions (left panel) and the two unresolved conditions (right panel), as a function of the overall years of

musical training. The duration of training seemed to only partly explain some of the variance in the pitch-discrimination performance for the resolved conditions ($R^2 = 0.18$; $p = 0.103$).

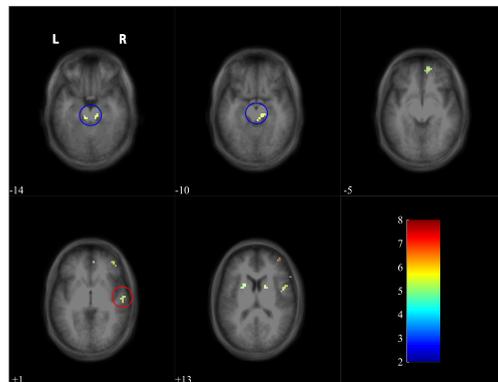
5.3.2 Experiment II: Cortical responses during a pitch-discrimination task

A full-factorial ANOVA, with the F_0 DLs and the behavioral performance entered as regressors, revealed a significant effect of musical training on the cortical neural responses to pitch stimuli. All significant clusters of activation with at least 15 suprathreshold voxels are listed in Table 5.1 for the contrasts musicians>non-musicians and non-musicians>musicians. Figure 5.5 depicts the differential activation map of musicians relative to non-musicians in response to all pitch conditions *vs.* noise, superimposed to the mean T1-weighted image ($p < 0.05$ FWE). Areas of increased activation in musicians comprised the inferior colliculus (blue circles in Fig. 5.5), the right primary auditory cortex (Heschl's gyrus; red circles in Fig. 5.5), the right middle (36, 44, 7; $t = 7.97$) and right superior (24, 56, 10; $t = 5.43$) frontal gyri, the right insula (42, -1, 16; $t = 5.83$) and frontal operculum (54, 8, 19; $t = 6.04$). Thus, although the task difficulty was adjusted across participants, the BOLD response for the musically trained listeners was significantly larger than for the non-musicians, with a stronger right-lateralized activation in the primary auditory cortex and frontal regions. Additionally, the analysis revealed a significantly larger BOLD signal in the non-musicians relative to the musicians in the right hippocampus (33, -43, 7; $t = 6.91$). A stronger activation of the right hippocampus in the non-musicians might be related to an increased allocation of short-term memory for pitch retrieval (Fortin et al., 2002; Lehn et al., 2009).

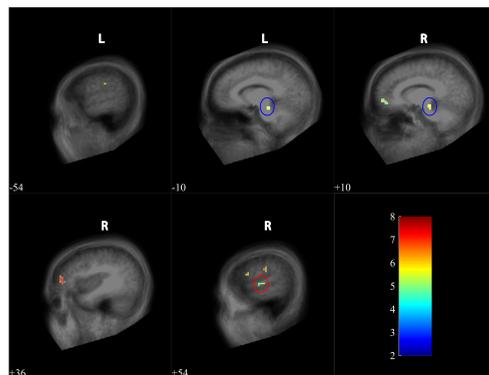
Furthermore, the ANOVA revealed a significant effect of task difficulty (Fig. 5.6a, $p < 0.001$, uncorrected; Table 5.1, $p < 0.05$, FWE), with the difficult conditions tested at 60% showing a larger activation relative to the easy conditions tested at 90%. This effect was significant in the right middle (30, 41, 10; $t = 4.76$) and right inferior (pars triangularis: 39, 20, 13; $t = 4.07$) frontal gyri, left putamen (-24, 8, 16; $t = 4.82$) and left caudate nucleus (-21, 14, 16; $t = 4.57$). Thus, the larger the difficulty imposed by



(a) Coronal



(b) Axial



(c) Sagittal

Figure 5.5: Differential activation map to the contrast musicians > non-musicians, superimposed to the mean T1-weighted image ($p < 0.05$, FWE). From top to bottom: coronal, axial and sagittal slices. The right auditory cortex is highlight by a red circle and the inferior colliculus by a blue circle. The color scale refers to the t-values.

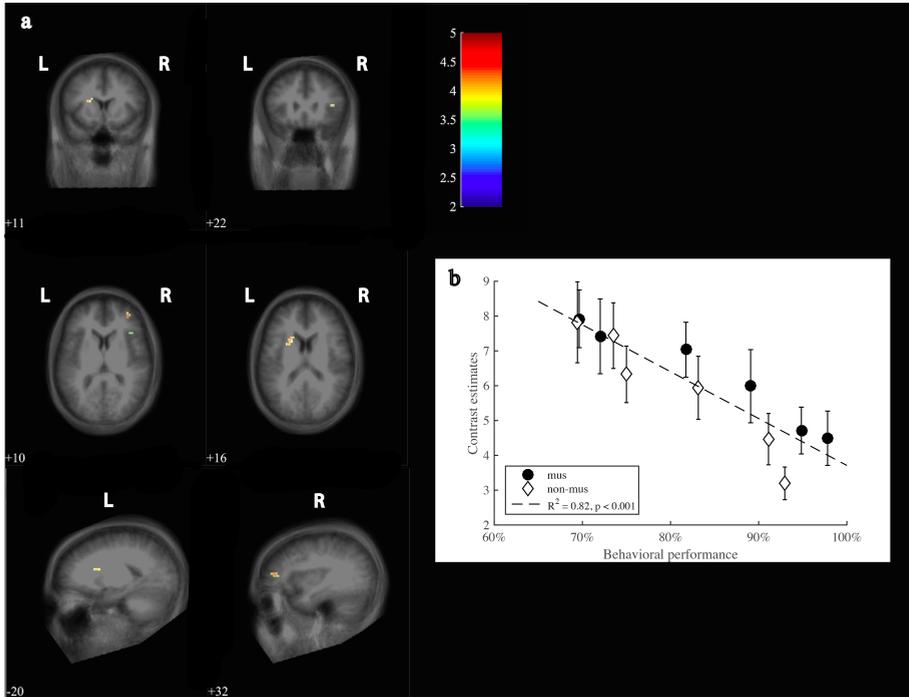


Figure 5.6: a) Differential activation map to the contrast 60%>90%, superimposed to the mean T1-weighted image ($p < 0.001$, uncorr.). From top to bottom: coronal, axial and sagittal slices. The color scale refers to the t-values. b) Correlation between the mean activation in the frontal mask defined in Fig.5.3(b) and the behavioral performance (% correct deviant identification) for the six tested conditions. Mean across the 15 musicians (closed circles) and the 15 non-musicians (open diamonds).

the pitch-discrimination task, the larger was the activation of these frontal regions for both musicians and non-musicians, suggesting an increase in working memory and effort with increasing task difficulty (Zatorre et al., 1994; Albouy et al., 2013). Additionally, the mean activation within the frontal mask defined in Fig.5.3(b) was calculated for each pitch condition (relative to noise) and each participant. The increase in the mean activation across participants was significantly correlated with a decrease in the behavioral performance for both groups of listeners (Fig. 5.6b).

Finally, the ANOVA did not reveal any larger neural activation for the resolved

Table 5.1: Full-factorial ANOVA ($p < 0.05$ FWE): effects of musicianship (only clusters with more than 15 suprathreshold voxels are reported) and task difficulty (difficult task > easy task).

Region	Voxels per cluster	Coordinates			t-value
		x	y	z	
musicians > non-musicians					
Inferior colliculus	141	9	-25	-11	6.18
R Mid/Sup Frontal Gyrus	113	36	44	7	7.97
R Insula/Frontal Operculum	88	54	8	19	6.04
R Mid Orbital Gyrus	85	12	56	-2	5.97
R Heschl's Gyrus (Te1.0, Te1.1); Te3	78	63	-10	1	6.05
R Rolandic Operculum	64	54	-22	25	6.79
L Caudate/Putamen	58	-24	5	16	6.87
L Parietal Operculum	41	-60	-25	28	6.78
R Caudate	34	12	5	16	5.82
Left Fusiform Gyrus	26	-27	-82	-5	5.73
R Cerebellum	20	18	-52	-35	5.78
L Heschl's Gyrus (Te1.2); Te3	19	-51	-13	7	5.08
L Thalamus	17	-3	-10	22	5.79
non-musicians > musicians					
R Hippocampus	47	33	-43	7	6.91
60% > 90%					
R Mid Frontal Gyrus	4	30	41	10	4.76
L Putamen	1	-24	8	16	4.82
L Caudate	1	-21	14	16	4.57

conditions relative to the unresolved conditions, in contrast to previous studies (Penagos et al., 2004; Norman-Haignere et al., 2013). To investigate this aspect further, a region of interest (ROI) analysis was carried out to identify the pitch-sensitive voxels within the right and left primary and secondary auditory cortices (Norman-Haignere et al., 2013). The 10% most activated voxels for all pitch conditions relative to the noise were selected in the primary and non-primary auditory cortices (i.e., in Te1.0, Te1.1, Te1.2 and Te3, see Fig. 5.3(a)) for each participant.

In an independent analysis, the response of these pitch-sensitive voxels was estimated for all six tested conditions (four resolved conditions and two unresolved conditions). No increase in neural activation was found with increasing harmonic resolvability. The lack of a resolvability effect is discussed further in the discussion section.

The mean activation of the pitch-sensitive voxels across conditions was calculated for each participant. The increase in activation of the pitch-sensitive voxels in the right auditory cortex was significantly correlated with a finer F_0 -discrimination ability in the musicians (Fig. 5.7, top left panel). Thus, the finer the pitch-discrimination performance obtained in Experiment I for the musically trained listeners, the stronger the neural activation of the right auditory cortex obtained from Experiment II. No correlation was found neither for the musicians in the left auditory cortex (Fig. 5.7, top right panel), nor for the non-musicians' in either right or left auditory cortices (Fig. 5.7, bottom panels).

5.4 Discussion

5.4.1 The musicians' benefit in pitch discrimination for resolved and unresolved harmonics

The results from Experiment I revealed that the musicians' enhancement in pitch-discrimination performance relative to the non-musicians was larger for the resolved conditions, on average by about a factor of three, than for the unresolved conditions (factor of about two). A possible explanation for this finding is that natural musical sounds are harmonic complex tones that contain both resolved and unresolved harmonics. Thus, musicians are neither exposed nor trained to retrieve the pitch of complex tones containing only unresolved harmonics. Despite not being specifically trained on these stimuli, the musicians still showed better pitch-discrimination performance than the non-musicians for the unresolved conditions. This finding is in agreement with previous studies showing that learning is only partly resolvability-specific (Grimault et al., 2002; Carcagno and Plack, 2011). In fact, listeners trained with a resolved complex tone showed larger improvements

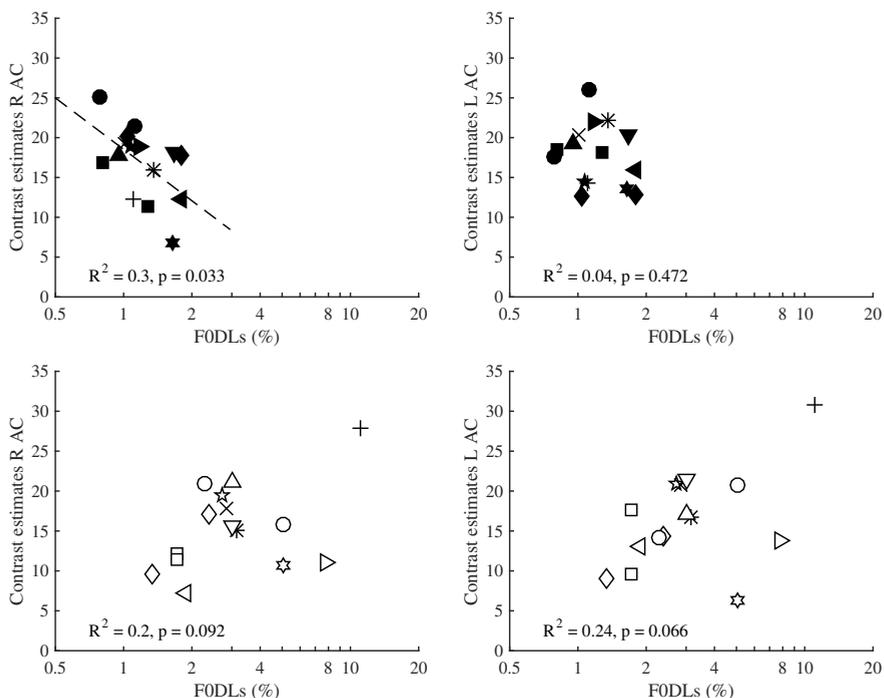


Figure 5.7: Individual contrast estimates (pitch > noise) in the right (left panels) and left (right panels) auditory cortices as a function of the individual pitch-discrimination thresholds from Experiment I, averaged across all six conditions. Values on the ordinate depict the mean of the 10% voxels with the highest activation within the mask of Fig. 5.3(a). Top panels: 15 musicians (filled symbols; same symbols as in Fig. 5.4(b)); bottom panels: 15 non-musicians (open symbols).

in pitch-discrimination performance for another resolved complex tone than for an unresolved condition (Grimault et al., 2002). Thus, although learning generalized to the untrained condition, the transfer of learning to a stimulus with a different resolvability was not complete. Similarly, Micheyl et al. (2006) showed that the musicians' advantage in pitch discrimination over non-musicians was larger for complex tones (with resolved harmonics) than for pure tones, consistent with an incomplete generalization of learning with unfamiliar and unnatural sounds (Demany and Semal, 2002).

In Bianchi et al. (2016b), the increase in pitch-discrimination performance with musical training was similar (a factor of about two) for complex tones with either resolved or unresolved harmonics. In the present study, the musicians' benefit was larger for the resolved complex tones, i.e., a factor of three relative to the non-musicians. There are two possible explanations for this discrepancy in the results. First, it could be due to the stricter musicians' inclusion criterion used in the present study (at least 8 years of formal musical education), although there was only a mild tendency of decreasing F_0 DLs with increasing years of overall musical training for the resolved conditions (see Fig. 5.4(c), left panel). The second explanation could be the diverse distribution of the musicians across the families of played instruments (see Fig. 5.4(b)). In fact, in the present study, most musicians were typically required to tune their own instrument before playing (e.g., string instruments), or to match the pitch of their voice to the accompanying instrument or to the other singers. Differently, only two musicians tested in Bianchi et al. (2016b) were string players. Musicians who need to tune their instruments have, indeed, been found to be more sensitive to fine pitch changes relative to musicians who do not tune their instrument themselves (e.g., Spiegel and Watson, 1984; Micheyl et al., 2006). In the current study, string players, singers and classical percussionists were the most sensitive to pitch-discrimination changes, especially for resolved harmonics (see Fig. 5.8), while piano players were the least sensitive. No conclusion on the two woodwinds players can be drawn since they both stopped playing regularly, which might have influenced their pitch-discrimination abilities. It should also be noted that the percussionists might have obtained lower thresholds due to the genre of played music (classical) rather than the played instrument per se (Micheyl et al., 2006). Interestingly, the musicians who performed the best in pitch discrimination of unresolved complex tones (Fig. 5.8) were studying to be professional singers. An explanation for their enhanced pitch-discrimination performance with unresolved complex tones relative to the other musicians might be that the singers are more exposed to unresolved harmonics, since the vocal tract resonance frequencies produce a considerable enhancement of harmonics in a frequency region between 2 and 4 kHz (Sundberg, 1977; Wolfe et al., 2009). Thus, at relatively low F_0 s (e.g., alto or tenor voice: $F_0 = 220$ Hz), high-numbered unresolved harmonics are enhanced

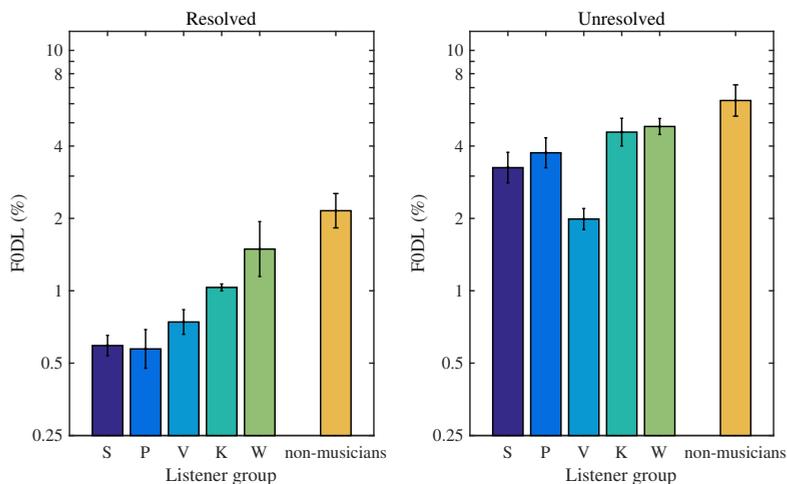


Figure 5.8: Mean pitch-discrimination thresholds for the four resolved conditions (left panels) and the two unresolved conditions (right panels) for the 16 musicians, divided by the family of their main instrument (S: string; P: percussions; V: voice; K: keyboard; W: woodwinds) and the 15 non-musicians.

(e.g., harmonics from the 9th to the 18th for the alto or tenor voice). Overall, the findings from Experiment I suggested that the musicians' benefit in pitch-discrimination performance was highly dependent on the family of played instruments and differed for stimuli containing only resolved or unresolved harmonics.

5.4.2 The musicians' benefit and implications for a post-peripheral advantage

While some studies have suggested that experience-dependent changes in musicians emerge already at the level of the cochlea in terms of sharper peripheral tuning (Soderquist, 1970; Bidelman et al., 2014; Bidelman et al., 2016), other studies did not find evidence of narrower peripheral auditory filters in musicians (Fine and Moore, 1993; Oxenham et al., 2003; Bianchi et al., 2016b). If the musicians' benefit observed in this study was ascribed to sharper peripheral frequency selectivity, an advantage in pitch-discrimination would be expected only for resolved complex tones, while no advantage would be expected to occur for the unresolved complex

tones. In fact, narrower peripheral auditory filters were found to enhance the pitch-discrimination performance of resolved complex tones and to shift the F_0 transition point, around which the harmonics become resolved, towards smaller F_0 s (e.g., Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b). Moreover, no systematic increase in performance was observed for listeners with narrower peripheral auditory filters for pitch discrimination of unresolved complex tones with harmonics added in sine phase (e.g., Bernstein and Oxenham, 2006a; Bernstein and Oxenham, 2006b; Bianchi et al., 2016a). A decrease in auditory filter bandwidth would in fact reduce, and not enhance, the temporal envelope cues available at the output of the filter, due to fewer harmonics interacting within the same filter. Thus, the musicians' benefit obtained in the present study for pitch discrimination of peripherally unresolved harmonics cannot be accounted for by narrower peripheral auditory filters. However, an enhanced temporal coding (e.g., enhanced neural synchrony; Parbery-Clark et al., 2009) at the level of the auditory nerve might still occur and enhance the envelope representation for unresolved stimuli, as well as fine structure cues. The behavioral findings of the current study suggest that an enhanced F_0 representation occurs along the auditory system in musicians at stages beyond the cochlea and applies to both resolved and unresolved complex tones, with a lower benefit for the unresolved harmonics, possibly as a result of the lack of familiarity with these stimuli.

5.4.3 Neural correlates of resolvability

Two previous studies have reported an effect of harmonic resolvability in anterior regions of the auditory cortex, where complex tones with resolved harmonics elicited stronger responses compared to complex tones containing only unresolved harmonics (Penagos et al., 2004; Norman-Haignere et al., 2013). This effect of resolvability was not observed in the current study. In the following paragraphs, possible reasons for the absence of this effect are discussed.

First, in the current stimulus design, the level per harmonic was fixed, leading to the same signal to noise ratio (SNR) per harmonic in all conditions relative to the noise, but to a higher overall stimulus level for the unresolved conditions than

for the resolved conditions. In previous studies (e.g., Penagos et al., 2004; Norman-Haignere et al., 2013), the overall sound pressure level was, instead, kept constant independent of the number of harmonics present in the stimulus. Thus, the SNR of each harmonic for a resolved complex tone containing only few harmonics was much higher than the SNR per harmonic for an unresolved complex tone with four times more components (e.g., Penagos et al., 2004; Norman-Haignere et al., 2013). Thus, the increase in neural activation with increasing resolvability found in these studies might have also reflected an increase in the SNR of each harmonic. This is plausible considering that the SNR-sensitive region in the auditory cortex was found to largely overlap with the pitch-sensitive region (Ernst et al., 2008). On the other hand, the overall sound pressure level of the unresolved conditions used in the current study was higher than that of the resolved conditions with few harmonics (i.e., the conditions at 500-Hz F_0). As correlates of overall level have also been reported in the primary auditory cortex and planum temporale (Ernst et al., 2008), it is possible that the effect of resolvability might have been counteracted by changes in overall level.

The second possible explanation for the lack of a resolvability effect is that the F_0 between reference and deviant was adjusted according to the individual F_0 DLs obtained in Experiment I. Thus, the unresolved conditions were tested at a larger F_0 between reference and deviant than the resolved conditions, which might have led to a similar pitch salience across conditions and, thus, to similar neural responses. In favor of this hypothesis, increasing the pitch interval size during a melody-discrimination task (Zatorre et al., 2012) or during passive listening of pure-tone melodies (Hyde et al., 2008) has been shown to increase the neural activation in the auditory cortex. However, in Zatorre et al. (2012), increasing the frequency separation between notes increased the behavioral performance, whereas in the current study the F_0 was increased to match the behavioral performance across conditions and participants.

Finally, although a TEN at 45 dB/ERB was used to mask distortion products (DPs) and the stimulus level was relatively moderate at 50 dB SPL per harmonic, it is possible that either cochlear or earphone distortions introduced audible low-numbered resolved harmonics (Norman-Haignere and McDermott, 2016). However,

cochlear and earphones DPs for harmonic levels of 70 dB SPL introduced low-frequency harmonic components at levels of about 30 to 50 dB SPL (Norman-Haignere and McDermott, 2016). Thus, for the stimuli of the current study, DPs should be below 30 dB SPL, and should have, therefore, been masked by the TEN at 45 dB SPL/ERB.

5.4.4 Correlates of musical training in the right primary auditory cortex

While the behavioral results from Experiment I revealed a stronger benefit of musicians for pitch discrimination of resolved relative to unresolved complex tones, the fMRI results confirmed enhanced neural responses in the musicians for both resolved and unresolved conditions in the right Heschl's gyrus, right superior and middle frontal gyri, right insula, right frontal operculum and inferior colliculus (Fig. 5.5). Additionally, the increase in neural activation in the right auditory cortex was significantly correlated with the increase in pitch-discrimination performance for the musically-trained listeners (Fig. 5.7). Taken altogether, these results are consistent with a right lateralization of cortical processing during a pitch-discrimination task involving fine pitch differences (e.g., Zatorre, 1988; Johnsrude et al., 2000; Zatorre and Belin, 2001; Zatorre et al., 2002; Hyde et al., 2008). Additionally, these findings suggest an increasing training-dependent plasticity in the right auditory cortex with increasing pitch-discrimination aptitude (Schneider et al., 2002; Puschmann et al., 2013).

A further analysis revealed that the correlation between BOLD responses in the auditory cortex and individual pitch-discrimination thresholds was restricted to the resolved conditions (Fig. 5.9). Figure 5.9 depicts the relation between the neural responses to the contrast pitch>noise in the right (left panels) and left (right panels) auditory cortex and the F_0 DLs for the 15 musicians. The contrast estimates for the two resolved conditions in the LF region are shown in the top panels, the estimates for the two resolved conditions in the HF region are depicted in the middle panels, while the two unresolved conditions are presented in the bottom panels. A strong significant correlation was obtained only in the right auditory

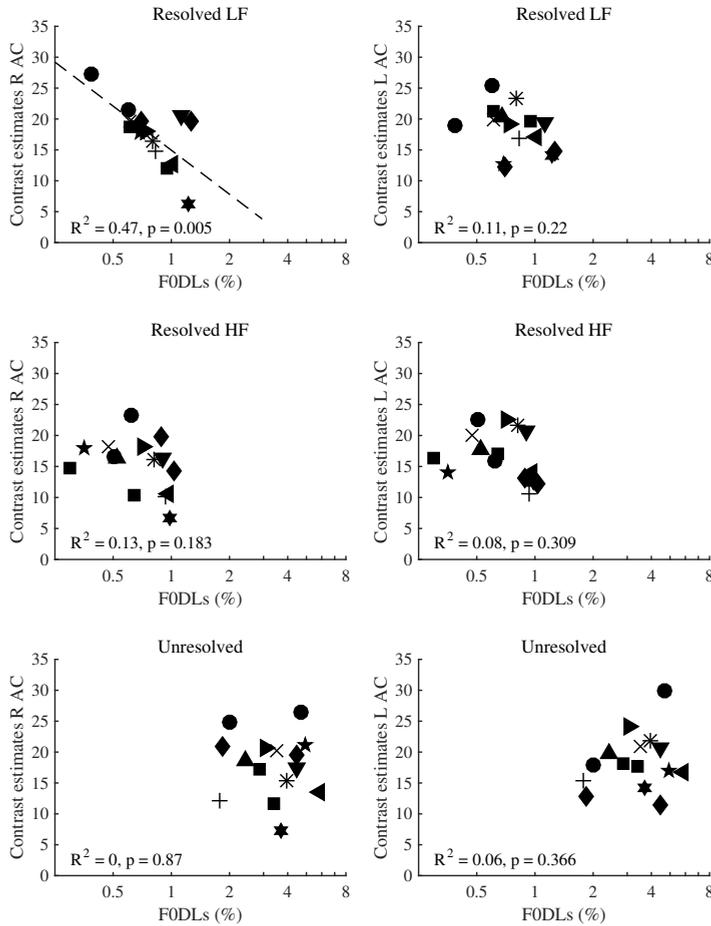


Figure 5.9: Individual contrast estimates (pitch > noise) in the right (left panels) and left (right panels) auditory cortices as a function of the individual pitch-discrimination thresholds from Experiment I, averaged across the two LF conditions (top panels), the two resolved HF conditions (middle panels) and the two unresolved conditions (bottom panels). Results for the 15 musicians (same symbols as in Figure 4). The individual contrast estimates were obtained by averaging the BOLD signal within the 10% most activated voxels in the right and left auditory cortices, as defined by the anatomical mask in Fig. 5.3(a).

cortex for the resolved conditions in the LF regions (top left panel in Fig. 5.9). A possible explanation for this finding is that the musicians, differently than the singers, might be more trained to retrieve the pitch from harmonics in a LF region than in a HF region, since the spectral envelope of the sound from a symphony orchestra produces a gradual decay in the amplitude of harmonics above about 1 kHz (Sundberg, 1977). Thus, a training-dependent plasticity might be stronger for the familiar LF-filtered stimuli than for the unfamiliar HF-filtered complex tones. Finally, it should be noted that the musicians' neural responses while performing a pitch-discrimination task were enhanced already at a subcortical level (inferior colliculus, Fig. 5.5). Enhanced subcortical responses in musicians were previously reported in electrophysiological studies (Musacchia et al., 2007; Wong et al., 2007; Parbery-Clark et al., 2009; Anderson and Kraus, 2011) and suggest that a training-dependent effects in the musicians may originate earlier in the auditory system than the auditory cortex.

5.5 Conclusion

Overall, these findings suggest an involvement of fronto-temporal regions comprising primary and non-primary auditory cortices, middle and superior frontal gyri, insula and frontal operculum during a pitch-discrimination task with conditions of varying task difficulty. When the harmonic level was fixed above the noise and the difference in F_0 was individually adjusted, no effect of harmonic resolvability was observed. Cortical responses to pitch in musicians were enhanced relative to non-musicians in the right Heschl's gyrus, right insula, right middle and superior frontal gyri. Interestingly, BOLD responses in the right auditory cortex were predictive of individual pitch-discrimination abilities in musicians, consistent with a higher specialization of the right auditory cortex in processing fine pitch changes relative to the left auditory cortex. Additionally, neural activation in the inferior colliculus was larger in the musicians than in the non-musicians, suggesting an increased pitch sensitivity already at a subcortical level. Finally, the increase in activation of frontal regions comprising inferior frontal gyrus, insula, and frontal

operculum was correlated with the decrease in behavioral performance, suggesting an increase of task-related effort and working memory with increasing the task difficulty for both musicians and non-musicians.

Acknowledgments

This work was supported by the Technical University of Denmark, Oticon Foundation and a grant from the Erasmus Mundus in Auditory Cognitive Neuroscience.

6

General discussion

In this thesis, behavioral investigations of pitch discrimination were carried out in normal-hearing (NH) and hearing-impaired (HI) listeners, as well as in musicians and non-musicians. Additionally, objective correlates of harmonic resolvability, task difficulty and musical training were investigated via pupillometry and functional magnetic resonance imaging (fMRI) to clarify the origin of the musicians' pitch-discrimination advantage. Differences in performance across the groups of listeners were discussed in relation to the nature of relevant cues for the human auditory system to retrieve pitch. Open questions, such as the enhancement of temporal envelope cues in hearing-impaired listeners and the effects of musical training on pitch discrimination of resolved and unresolved complex tones, were addressed throughout the chapters of this thesis.

6.1 Summary and discussion of main results

In the following paragraphs, the main outcomes of each chapter are summarized and discussed in relation to the main research questions addressed in this thesis.

6.1.1 Pitch discrimination performance of hearing-impaired listeners: Importance of envelope cues

In Chapter 2, the pitch-discrimination performance of NH listeners and listeners with sensorineural hearing loss (SNHL) was estimated for resolved and unresolved complex tones. When the harmonics were assumed to be unresolved for both group of listeners (i.e., when the lowest harmonic number was either 12 or 15), the performance of the HI listeners for complex tones with harmonics added in random

phase (RP) was significantly worse than that of the NH listeners. However, when the harmonics were added in sine phase (SP), the HI listeners' performance was as accurate as that of the NH listeners. Since the interaction between harmonics added in SP leads to a peakier envelope at the output of the auditory filters than for harmonics added in RP, the difference in performance between the RP and the SP condition was considered as an indicator of temporal envelope processing. The HI listeners showed a larger increase in pitch-discrimination performance for the SP relative to the RP condition as compared to the NH listeners. It was hypothesized that this larger benefit in performance for the SP relative to the RP condition might be explained by an enhancement of the internal envelope representation of a SP-complex tone in listeners with SNHL. The hypothesis that this enhancement might be related to cochlear damage was tested by estimating cochlear compression and auditory-filter bandwidths in the same listeners. A significant correlation was found between the reduction of cochlear compression and the increase in pitch-discrimination performance for the SP relative to the RP condition. Additionally, the relative effects of cochlear compression and auditory-filter bandwidth on the internal envelope representation of unresolved complex tones were considered in a simplified peripheral model. The model revealed that loss of cochlear compression was the dominant factor in enhancing the envelope peakiness of the SP, but not RP unresolved complex tones. Finally, the difference in the predicted envelope peakiness between SP and RP could account for the behavioral increase in pitch-discrimination performance of the SP relative to the RP condition.

Overall, the findings of this study suggest that a reduction of cochlear compression following SNHL may lead to enhanced temporal envelope cues for unresolved complex tones with harmonics added in SP. In agreement with this hypothesis, the HI listeners showed a higher sensitivity to detect amplitude modulations imposed on a sinusoidal carrier as compared to NH listeners, consistent with previous findings (Moore et al., 1996; Moore and Glasberg, 2001). However, the HI listeners did not show, on average, a better pitch-discrimination performance relative to NH listeners. There are different explanations for the lack of an enhanced pitch-discrimination performance. First of all, differences in musical education between

NH and HI listeners might have influenced the absolute levels of performance. In fact, although in each group of listeners 40% of the participants were musically trained (six out of 14 NH listeners; four out of 10 HI listeners), the overall duration of training as well as the families of played instruments were not matched across groups. Additionally, the HI listeners were much older than the NH listeners. This may lead to higher levels of internal noise, as well as to limitations in cognitive resources. Finally, other limitations linked to SNHL (e.g., inner hair cell damage, degradation of auditory-nerve coding) might have counteracted the enhanced envelope representation at the output of peripheral stages of the auditory system.

An enhanced temporal envelope coding has so far only been reported from physiological recordings in auditory-nerve fibers of chinchillas with noise-induced SNHL (Kale and Heinz, 2010; Kale and Heinz, 2012; Henry et al., 2014), and suggested by a few behavioral studies in human listeners (Moore et al., 1996; Moore and Glasberg, 2001). The present study demonstrated for the first time that the enhanced envelope representation at the output of peripheral filters for complex tones with unresolved harmonics added in SP was correlated to the reduction of cochlear compression in listeners with SNHL.

6.1.2 Pitch discrimination performance of musicians vs non-musicians: Effort and plasticity

In Chapter 3, pupil responses were recorded while musicians and non-musicians performed a pitch-discrimination task with resolved and unresolved complex tones. The aim of this study was to clarify whether the increase in pupil size could reflect the processing effort of the listeners while performing a pitch-discrimination task with stimuli of decreasing harmonic resolvability (i.e., decreasing pitch salience) and increasing task difficulty. The results showed that pupil size increased with concomitantly decreasing harmonic resolvability and increasing task difficulty for the musically-trained listeners. In contrast, the non-musicians showed a decrease in pupil size for the most demanding condition (unresolved harmonics tested with a difficult task). This finding, together with the low behavioral performance obtained for this condition (42% correct), suggests that the non-musicians might

have "given up" to perform a task beyond their discrimination abilities. A decrease in pupil size for highly demanding tasks was previously interpreted as a cognitive processing overload (Zekveld and Kramer, 2014).

Although it could be tempting to conclude that the musicians put more effort in discriminating the pitch of the most demanding condition, it should be considered that there was a large difference in the behavioral performance between the two groups of listeners (42% *vs.* 63%). Since the task appeared not to be equally demanding across participants, effort could not be directly compared across the two groups of listeners. In this study, the task difficulty was individually adjusted according to the behavioral thresholds for only six out of eleven participants. Additionally, for these six participants, only the task difficulty of the $F_{0,lr}$ point was adjusted. Thus, most conditions were not matched in task difficulty across participants. As a result, some conditions resulted to be very easy for the musicians (99% of correct deviant identification) and some conditions too difficult for the non-musicians (42% correct). A comparison of the processing effort allocated by the musicians relative to the non-musicians for either resolved or unresolved complex tones was, thus, confounded by the different levels of task difficulty.

Overall, the findings of this study revealed that it was possible to measure processing effort during a pitch-discrimination task via pupillometry. However, it remained unclear whether this change was driven by a change in the stimulus salience (i.e., in the resolvability of the harmonics) and/or by a change in task difficulty. Furthermore, it could not be clarified whether musicians allocate a different processing effort than non-musicians for complex-tone pitch discrimination. These open questions led to a new experimental design, presented in Chapter 4, which allowed for matching the difficulty of the task across conditions and participants and to disentangle the effects of harmonic resolvability and task difficulty on pupil dilations.

In Chapter 4, the musicians' benefit in pitch discrimination relative to non-musicians was first estimated behaviorally for resolved and unresolved complex tones. The musicians' increase in performance was of about a factor of two (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001), independent of harmonic resolvability. Additionally, the transition point from unresolved to resolved harmonics

occurred at similar F_0 s for musicians and non-musicians. Hence, it was suggested that the musicians' benefit in pitch discrimination may not originate peripherally, as reported in two recent studies (Bidelman et al., 2014; Bidelman et al., 2016). In fact, narrower peripheral auditory filters in musicians should have shifted the transition point at lower F_0 s, leading to resolved harmonics at smaller F_0 s than for non-musicians (i.e., at higher harmonic numbers). Additionally, narrower peripheral auditory filters in musicians are not consistent with the behavioral enhancement obtained in this study for unresolved complex tones. In fact, the benefit in the musicians' performance for complex tones with only unresolved harmonics cannot be explained based on a finer peripheral frequency selectivity. Overall, the behavioral findings of Chapter 4 suggest a training-dependent benefit in musicians for both resolved and unresolved complex tones, which is plausible to occur *after* peripheral stages of the auditory system. It seems likely that a training-dependent effect in musicians occurs after F_0 extraction, possibly at central stages of the auditory system.

After ruling out the possibility of a peripheral enhancement in musicians, Chapter 4 addressed the question on whether musicians may allocate a different processing effort to perform a pitch-discrimination task with resolved and unresolved complex tones. Although the behavioral benefit of musicians was similar for the two types of stimuli, it may still have different origins. In fact, one might speculate that since musicians are mostly exposed to natural harmonic complex sounds containing all harmonics, a training-dependent effect may be specific to the low-numbered harmonics, which are assumed to be dominant for pitch perception (Plack, 2005). This hypothesis was tested by presenting the listeners with a pitch-discrimination task with either resolved or unresolved harmonics and by recording pupil dilation, when the task difficulty was individually adjusted according to the behavioral results. The pupil size showed a significant increase with increasing task difficulty. Additionally, the interaction of task difficulty and harmonic resolvability was significant. Hence, pupil responses seemed to reflect the overall processing demand given by the interaction of both factors, whereby higher task difficulty and lower harmonic resolvability led to an increased processing effort. Additionally, pupil dilations for musicians were smaller than dilations for non-musicians,

specifically when at least some of the harmonics were resolved and the task was either medium-difficult or easy (i.e., for conditions with low processing demand). Thus, although the task difficulty was adjusted to compensate for the musicians' finer pitch-discrimination abilities, the musically trained listeners still allocated lower processing effort than did the non-musicians to perform the task at the same performance level. This finding suggests an enhanced pitch representation along the auditory system in musicians, possibly as a result of training, which seemed to be specific to the stimuli containing resolved harmonics (or general to the stimuli with low processing demand).

The hypothesis of an enhanced pitch representation in musicians to resolved and unresolved complex tones was tested in Chapter 5, where cortical neural responses were investigated via fMRI. In this study, highly trained musicians and non-musicians first participated in a behavioral pitch-discrimination task. The results showed that the musicians' enhancement in pitch-discrimination performance relative to the non-musicians was larger for the resolved conditions, on average by about a factor of three, than for the unresolved conditions (factor of about two, similar as in Chapter 4). The increased benefit obtained in this study for the resolved complex tones might be ascribed to the stricter inclusion criterion of the musicians (relative to the study of Chapter 4). This argument further supports the hypothesis of a training-dependent effect specific to the resolved harmonics. The fMRI results showed enhanced neural responses in the musicians relative to the non-musicians for both resolved and unresolved complex tones. This enhancement was specific to the right Heschl's gyrus, right superior and middle frontal gyri, right insula, right frontal operculum and inferior colliculus. Additionally, the increase in neural activation in the right auditory cortex of musicians was predictive of the increased pitch-discrimination performance for resolved complex tones. These results suggest a training-dependent plasticity in the right auditory cortex of musicians, which seems to be specific to the resolved complex tones. In fact, although enhanced neural responses in musicians were also observed for the unresolved complex tones, no correlation was found with the behavioral pitch-discrimination performance. Additionally, the fMRI results obtained in this study revealed enhanced subcortical responses in musicians for both resolved and

unresolved complex tones, in agreement with previous electrophysiological studies (Musacchia et al., 2007; Wong et al., 2007; Parbery-Clark et al., 2009; Anderson and Kraus, 2011) and suggest that training-dependent effects in the musicians may originate earlier in the auditory system than the auditory cortex.

6.2 Implications for pitch coding mechanisms

While frequency selectivity and temporal fine structure (TFS) cues are known to be degraded in listeners with SNHL (Glasberg and Moore, 1986; Moore et al., 2006b; Lorenzi et al., 2006; Hopkins and Moore, 2007), the outcomes of the behavioral experiments presented in Chapter 2, together with the model predictions, suggest that the internal envelope representation of complex tones with harmonics added in SP is enhanced by the reduced cochlear compression. Thus, the relative importance of pitch cues in HI listeners seems to be altered relative to NH listeners (Kale and Heinz, 2010). In fact, while NH listeners were found to rely on resolvability and/or TFS cues for complex tones with intermediate harmonic numbers (e.g., when the lowest harmonic number was in the range 8-11; Moore and Moore, 2003; Moore et al., 2006b), HI listeners could only rely on envelope cues and their performance was decreased relative to NH listeners for harmonics in this range. However, when both NH and HI listeners were assumed to rely on temporal envelope cues, the performance of the two groups of listeners was found to be similar for the SP complex tones, suggesting that an enhanced envelope representation in HI listeners may be counteracted by other factors.

It should be noted that, in the model, an instantaneous compression was applied to the signal at the output of the filter, i.e., the compression worked on the TFS. Hence, although the modulation power was calculated based on the low-pass filtered envelope, the effects of (reduced) compression would be equally reflected in the TFS. Thus, the hypothesis that the increase in modulation power reflects the enhancement of temporal envelope cues, and not TFS cues, was based on the assumption that HI listeners can only rely on envelope cues to retrieve the pitch of such high-numbered harmonics (Moore and Moore, 2003; Moore et al., 2006b;

Hopkins and Moore, 2007). However, it cannot be excluded that NH listeners may have relied on TFS cues when the lowest harmonic number was 12 (Boer, 1956; Schouten et al., 1962; Moore and Moore, 2003; Moore and Sek, 2009; Kale et al., 2014).

In Chapter 4, the benefit of the musicians in pitch-discrimination was similar for resolved and unresolved complex tones. The interpretation of these results in terms of pitch coding mechanisms is not straightforward. First of all, the similar transition points in the thresholds from unresolved to resolved harmonics suggest an enhancement in performance that is unlikely to originate peripherally. Secondly, the almost identical size of the musicians' benefit for resolved and unresolved harmonics suggests either a training-dependent enhancement in musicians that occurs independently of resolvability (e.g., an enhanced neural synchrony after F_0 extraction; e.g., Parbery-Clark et al., 2009; Anderson and Kraus, 2011), or different physiological mechanisms leading to similar benefits. The fact that the musicians allocated lower processing effort than the non-musicians during the pitch-discrimination task with resolved harmonics, but not with unresolved harmonics, may suggest a training-dependent enhancement specific to the low-numbered harmonics. Thus, it may well be that the behavioral enhancement for the resolved complex tones reflects a training-dependent effect, while the behavioral enhancement for the unresolved complex tones is a consequence of the allocation of more processing effort to perform the task more accurately than the non-musicians.

In Chapter 5, the behavioral benefit of musicians in pitch discrimination was larger for the resolved than for the unresolved complex tones. Additionally, the high correlation between neural responses in the right auditory cortex and behavioral pitch-discrimination thresholds for the low-frequency resolved complex tones seemed to further support a training-dependent effect specific to the resolved harmonics. However, it is unclear whether these findings can be interpreted in terms of different mechanisms for pitch coding or simply support the hypothesis that the mechanisms for learning are only partly stimulus-specific (Carcagno and Plack, 2011). In fact, although the musicians are not specifically trained to discriminate the pitch of complex tones with only unresolved harmonics, they are in general more trained to perform a pitch-discrimination task. Thus, possibly

as a generalization of learning (Grimault et al., 2003), the musicians showed an enhanced pitch-discrimination performance for the unresolved complex tones, although this enhancement was not as strong as the one for the stimuli they are normally trained with (incomplete generalization of learning; Micheyl et al., 2006).

6.3 Perspectives

In conclusion, there are still many open questions to be addressed towards the understanding of the basic mechanisms underlying pitch perception in the human auditory system. Overall, the findings presented throughout this thesis strengthen the hypothesis that loss of cochlear compression leads to an enhanced envelope representation of unresolved complex tones in listeners with SNHL relative to NH listeners. However, a behavioral enhancement in HI listeners was observed only for amplitude-modulation detection and not for pitch-discrimination of unresolved complex tones, although both tasks should rely on envelope cues at low modulation/fundamental frequencies. Thus, although the findings of Chapter 2 suggest the presence of enhanced envelope cues in HI listeners, further objective measures should be carried out to confirm these results. Future work may focus on objective estimations of the envelope representation of complex stimuli in human listeners with SNHL (e.g., via frequency following responses).

It should be noted that while an enhancement of temporal envelope cues might be beneficial for pitch-discrimination of unresolved complex tones and amplitude-modulation detection, it could have a detrimental effect on speech intelligibility in fluctuating background noise. In fact, stronger envelope coding would enhance the fluctuations of the background noise, leading to a reduced ability of the HI listeners to listen in the dips of the masker (Festen and Plomp, 1990; Moore and Glasberg, 1993; Moore et al., 1995; Kale and Heinz, 2010). Thus, clarifying what factors lead to a stronger coding of envelope cues in HI listeners and understanding the extent of this enhancement would be necessary to restore normal envelope coding in the hearing-impaired population and to improve speech intelligibility in fluctuating background noise.

Additionally, both behavioral pitch-discrimination thresholds, as well as pupil responses and fMRI results suggest the existence of a training-dependent effect in musicians that is partly specific to the resolvability of the harmonics and may occur as early as the brainstem. Future research may clarify whether musical training leads to a finer F_0 representation for both resolved and unresolved harmonics. Tonotopic maps obtained with complex-tone stimuli could, in fact, reveal the effects of musical training on the harmonic organization of the auditory cortex. It would be interesting to clarify whether a finer spectral tuning may arise as a result of musical training. If so, enhancing cortical tuning by means of musical training may be beneficial for HI listeners, whereby musical training may help restoring the relative importance of resolved *vs.* unresolved harmonics for pitch perception, thus potentially improving sound source segregation.

Bibliography

- Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). "The variability of human, BOLD hemodynamic responses." In: *NeuroImage* 8.4, pp. 360–9.
- Albouy, P. et al. (2013). "Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex". In: *Brain* 136.5, pp. 1639–1661.
- Allen, E. J. and Oxenham, A. J. (2014). "Symmetric interactions and interference between pitch and timbre." In: *The Journal of the Acoustical Society of America* 135.3, pp. 1371–9.
- Anderson, S. and Kraus, N. (2011). "Neural Encoding of Speech and Music: Implications for Hearing Speech in Noise". In: *Seminars in Hearing* 32.2, pp. 129–141.
- Arehart, K. H. (1994). "Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners." In: *The Journal of the Acoustical Society of America* 95.6, pp. 3574–3585.
- Arehart, K. H. and Burns, E. M. (1999). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss". In: 106.2, pp. 993–997.
- Barker, D., Plack, C. J., and Hall, D. A. (2011). "Human auditory cortical responses to pitch and to pitch strength." In: *Neuroreport* 22.3, pp. 111–115.
- Barrett, K. C., Ashley, R., Strait, D. L., and Kraus, N. (2013). "Art and science: how musical training shapes the brain." In: *Frontiers in psychology* 4.October, pp. 1–13.
- Beatty, J. (1982). "Task-evoked pupillary responses, processing load, and the structure of processing resources". In: *Psychological Bulletin* 91, pp. 276–292.

- Bendor, D. and Wang, X. (2005). "The neuronal representation of pitch in primate auditory cortex". In: *Nature* 436.7054, pp. 1161–1165.
- Bermudez, P. and Zatorre, R. J. (2005). "Differences in gray matter between musicians and nonmusicians." In: *Annals of the New York Academy of Sciences* 1060.2005, pp. 395–399.
- Bernstein, J. G. W. and Oxenham, A. J. (2003). "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number?" In: *The Journal of the Acoustical Society of America* 113.6, pp. 3323–3334.
- Bernstein, J. G. W. and Oxenham, A. J. (2006a). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level". In: *The Journal of the Acoustical Society of America* 120.6, pp. 3916–3928.
- Bernstein, J. G. W. and Oxenham, A. J. (2006b). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss". In: *The Journal of the Acoustical Society of America* 120.6, pp. 3929–3945.
- Bernstein, J. G. W. and Oxenham, A. J. (2008). "Harmonic segregation through mistuning can improve fundamental frequency discrimination." In: *The Journal of the Acoustical Society of America* 124.3, pp. 1653–1667.
- Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (2014). "Objective correlates of pitch salience using pupillometry". In: *Proceedings of Forum Acusticum 2014*.
- Bianchi, F., Fereczkowski, M., Zaar, J., and Santurette, S. (2016a). "Complex-tone pitch discrimination in listeners with sensorineural hearing loss". In: *Trends in Hearing (in press)*.
- Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (2016b). "Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort". In: *Journal of the Association for Research in Otolaryngology* 17.1, pp. 69–79.
- Bidelman, G. M. and Krishnan, A. (2009). "Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem." In: *The Journal of neuroscience* 29.42, pp. 13165–13171.
- Bidelman, G. M., Schug, J. M., Jennings, S. G., and Bhagat, S. P. (2014). "Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians." In: *The Journal of the Acoustical Society of America* 136.1, EL 33–39.

- Bidelman, G. M., Nelms, C., and Bhagat, S. P. (2016). "Musical experience sharpens human cochlear tuning". In: *Hearing Research* 335, pp. 40–46.
- Blaser, E., Eglinton, L., Carter, A. S., and Kaldy, Z. (2014). "Pupillometry reveals a mechanism for the Autism Spectrum Disorder (ASD) advantage in visual tasks." In: *Scientific reports* 4, p. 4301.
- Boer, E. de (1956). "Pitch of Inharmonic Signals". In: *Nature* 178, pp. 535–536.
- Bradshaw, J. (1967). "Pupil Size as a Measure of Arousal during Information Processing". In: *Nature* 216, pp. 515–516.
- Carcagno, S. and Plack, C. J. (2011). "Pitch discrimination learning: specificity for pitch and harmonic resolvability, and electrophysiological correlates." In: *Journal of the Association for Research in Otolaryngology* 12.4, pp. 503–517.
- Carlyon, R. P. and Shackleton, T. M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" In: *The Journal of the Acoustical Society of America* 95.6, pp. 3541–3554.
- Cedolin, L. and Delgutte, B. (2005). "Representations of the pitch of complex tones in the auditory nerve". In: *Auditory signal processing: Physiology, Psychoacoustics, and Models*. Ed. by D Pressnitzer, A de Cheveigné, S. Mc Adams, and L. Collet. Springer, pp. 107–116.
- Cedolin, L. and Delgutte, B. (2010). "Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30.38, pp. 12712–12724.
- De Cheveigné, A (2005). "Pitch Perception Models". In: *Pitch: Neural Coding and Perception*. Ed. by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper. Springer. Chap. 6, pp. 169–233.
- Demany, L. and Semal, C. (2002). "Learning to perceive pitch differences." In: *The Journal of the Acoustical Society of America* 111.3, pp. 1377–1388.
- Ernst, S. M. A., Verhey, J. L., and Uppenkamp, S. (2008). "Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise". In: *NeuroImage* 43.2, pp. 321–328.
- Ewert, S. D. and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations." In: *Journal of the Acoustical Society of America* 108.3 Pt 1, pp. 1181–96.

- Fereczkowski, M. (2015). "Time-efficient behavioral estimates of cochlear compression". PhD thesis. Lyngby: Technical University of Denmark.
- Festen, J. M. and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing." In: *The Journal of the Acoustical Society of America* 88.4, pp. 1725–1736.
- Fine, P. and Moore, B. C. J. (1993). "Frequency analysis and musical ability". In: *Music Perception: An Interdisciplinary Journal* 11.1, pp. 39–54.
- Fishman, Y. I., Micheyl, C., and Steinschneider, M. (2013). "Neural Representation of Harmonic Complex Tones in Primary Auditory Cortex of the Awake Monkey". In: *The Journal of Neuroscience* 33.25, pp. 10312–10323.
- Fortin, N. J., Agster, K. L., and Eichenbaum, H. B. (2002). "Critical role of the hippocampus in memory for sequences of events." In: *Nature neuroscience* 5.5, pp. 458–62.
- Foster, N. E. V. and Zatorre, R. J. (2010). "Cortical structure predicts success in performing musical transformation judgments". In: *NeuroImage* 53.1, pp. 26–36.
- Garcia, D., Hall, D. A., and Plack, C. J. (2010). "The effect of stimulus context on pitch representations in the human auditory cortex". In: *NeuroImage* 51.2, pp. 808–816.
- Glasberg, B. R. and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments." In: *The Journal of the Acoustical Society of America* 79.4, pp. 1020–1033.
- Glasberg, B. R. and Moore, B. C. J. (1989). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech". In: *Scand Audiol Suppl* 32, pp. 1–25.
- Glasberg, B. R. and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data". In: *Hearing Research* 47, pp. 103–138.
- Gockel, H., Moore, B. C. J., and Carlyon, R. P. (2001). "Influence of rate of change of frequency on the overall pitch of frequency-modulated tones." In: *The Journal of the Acoustical Society of America* 109.2, pp. 701–712.
- Gockel, H. E., Carlyon, R. P., Mehta, A., and Plack, C. J. (2011). "The frequency following response (FFR) may reflect pitch-bearing information but is not a

- direct representation of pitch". In: *Journal of the Association for Research in Otolaryngology* 12.6, pp. 767–782.
- Goldstein, J. (1973). "An optimum processor theory for the central formation of the pitch of complex tones". In: *Journal of the Acoustical Society of America* 54, pp. 1496–1516.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., and Dykes, K. L. (1996). "Pupillary responses index cognitive resource limitations". In: *Psychophysiology* 33, pp. 457–461.
- Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (2002). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm." In: *Perception & psychophysics* 64.2, pp. 189–197.
- Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., and Collet, L. (2003). "Learning in discrimination of frequency or modulation rate: Generalization to fundamental frequency discrimination". In: *Hearing Research* 184.1-2, pp. 41–50.
- Hall, D. A. and Barker, D. (2012). "Coding of Basic Acoustical and Perceptual Components of Sound in Human Auditory Cortex". In: *The Human Auditory Cortex*. New York: Springer. Chap. 7, pp. 165–197.
- Hall, D. A. and Plack, C. J. (2009). "Pitch processing sites in the human auditory brain". In: *Cerebral Cortex* 19.3, pp. 576–585.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve". In: *Neural Comput* 13.10, pp. 2273–2316.
- Helmholtz, H. von (1877). *On the sensations of tone*. New York: Dover.
- Henry, K. S., Kale, S., and Heinz, M. G. (2014). "Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers." In: *Frontiers in systems neuroscience* 8.20, pp. 1–10.
- Herholz, S. C. and Zatorre, R. J. (2012). "Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure". In: *Neuron* 76.3, pp. 486–502.
- Herholz, S. C., Halpern, A. R., and Zatorre, R. J. (2012). "Neuronal correlates of perception, imagery, and memory for familiar tunes." In: *Journal of Cognitive Neuroscience* 24.6, pp. 1382–1397.

- Hess, E. H. and Polt, J. M. (1964). "Pupil Size in Relation to Mental Activity during Simple Problem-Solving". In: *Science* 143.3611, pp. 1190–1192.
- Hoekstra, A. (1979). "Frequency discrimination and frequency analysis in hearing". PhD thesis. Groningen: Institute of Audiology.
- Hoekstra, A. and Ritsma, R. J. (1977). "Perceptive hearing loss and frequency selectivity". In: *Psychophysics and Physiology of Hearing*. Ed. by E. Evans and J. Wilson. London: Academic, pp. 263–271.
- Hopkins, K. and Moore, B. C. J. (2007). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information." In: *The Journal of the Acoustical Society of America* 122.2, pp. 1055–1068.
- Houtsma, A. J. M. and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics". In: *Journal of the Acoustical Society of America* 87, pp. 304–310.
- Hyde, K. L. et al. (2009). "Musical training shapes structural brain development". In: *Journal of Neuroscience* 29.10, pp. 3019–3025.
- Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). "Evidence for the role of the right auditory cortex in fine pitch resolution". In: *Neuropsychologia* 46.2, pp. 632–639.
- Janisse, M. (1977). *Pupillometry: The psychology of the pupillary response*. Washington, DC: Hemisphere Publishing Corporation.
- Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (2000). "Functional specificity in the right human auditory cortex for perceiving pitch direction." In: *Brain : a journal of neurology* 123 (Pt 1, pp. 155–163.
- Johnsrude, I. S. and Rodd, J. M. (2015). "Factors that increase processing demands when listening to speech". In: *Neurobiology of Language*. Ed. by G. Hickok and S. L. Small. Elsevier Inc. Chap. 40, pp. 491–502.
- Kaernbach, C. (1991). "Simple adaptive testing with the weighted up-down method". In: *Perception & Psychophysics* 49.3, pp. 227–229.
- Kaernbach, C. and Bering, C. (2001). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics." In: *The Journal of the Acoustical Society of America* 110.2, pp. 1039–1048.

- Kahneman, D. and Beatty, J. (1967). "Pupillary responses in a pitch-discrimination task". In: *Perception & Psychophysics* 2.3, pp. 101–105.
- Kale, S. and Heinz, M. G. (2010). "Envelope coding in auditory nerve fibers following noise-induced hearing loss". In: *Journal of the Association for Research in Otolaryngology* 11.4, pp. 657–673.
- Kale, S. and Heinz, M. G. (2012). "Temporal modulation transfer functions measured from auditory-nerve responses following sensorineural hearing loss". In: *Hearing Research* 286.1-2, pp. 64–75.
- Kale, S., Micheyl, C., and Heinz, M. G. (2014). "Implications of within-fiber temporal coding for perceptual studies of F0 discrimination and discrimination of harmonic and inharmonic tone complexes". In: *Journal of the Association for Research in Otolaryngology* 15.3, pp. 465–482.
- Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (2001). "Pitch discrimination: are professional musicians better than non-musicians?" In: *Journal of basic and clinical physiology and pharmacology* 12.2 Suppl, pp. 125–143.
- Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., and Kramer, S. E. (2014). "The pupil response is sensitive to divided attention during speech processing". In: *Hearing Research* 312, pp. 114–120.
- Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers". In: *The Journal of the Acoustical Society of America* 108.2, p. 723.
- Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., and Gandour, J. T. (2012). "Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans." In: *Neuropsychologia* 50.12, pp. 2849–59.
- Lehn, H., Steffenach, H.-A., Strien, N. M. V., Veltman, D. J., Witter, M. P., and Håberg, A. K. (2009). "A Specific Role of the Human Hippocampus in Recall of Temporal Sequences". In: *The Journal of Neuroscience* 29.11, pp. 3475–3484.
- Licklider, J. (1951). "A duplex theory of pitch perception". In: *Experientia* 7, pp. 128–134.
- Licklider, J. (1959). "Three auditory theories". In: *A study of a Science*. Ed. by S. Koch. New York: Psychology. Chap. Vol. I. Pp. 41–144.

- Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (2003). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing." In: *The Journal of the Acoustical Society of America* 113.2, pp. 951–960.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.49, pp. 18866–18869.
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). "Influence of musical and psychoacoustical training on pitch discrimination". In: *Hearing Research* 219.1-2, pp. 36–47.
- Micheyl, C., Divis, K., Wroblewski, D. M., and Oxenham, A. J. (2010). "Does fundamental-frequency discrimination measure virtual pitch discrimination?" In: *The Journal of the Acoustical Society of America* 128.4, pp. 1930–1942.
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. fifth edit. London: Academic Press.
- Moore, B. C. J. and Carlyon, R. P. (2005). "Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users". In: *Pitch: Neural Coding and Perception*. Ed. by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper. Vol. 24. Springer, pp. 234–277.
- Moore, B. C. J. and Glasberg, B. R. (1988). "Pitch perception and phase sensitivity for subjects with unilateral and bilateral cochlear hearing impairments". In: *Clinical Audiology*. Ed. by A. Quaranta. Bari, Italy, pp. 104–109.
- Moore, B. C. J. and Glasberg, B. R. (1990). "Frequency selectivity in subjects with cochlear loss and its effects on pitch discrimination and phase sensitivity". In: *Advances in Audiology*. Ed. by F. Grandori, G. Cianfrone, and D. Kemp. Basel: Karger, pp. 187–200.
- Moore, B. C. J. and Glasberg, B. R. (1993). "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech." In: *The Journal of the Acoustical Society of America* 94.4, pp. 2050–2062.
- Moore, B. C. J. and Glasberg, B. R. (2001). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired

- listeners.” In: *The Journal of the Acoustical Society of America* 110.2, pp. 1067–1073.
- Moore, B. C. J. and Peters, R. W. (1992). “Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity.” In: *The Journal of the Acoustical Society of America* 91.5, pp. 2881–2893.
- Moore, B. C. J. and Sek, A. (2009). “Sensitivity of the human auditory system to temporal fine structure at high frequencies.” In: *The Journal of the Acoustical Society of America* 125.5, pp. 3186–3193.
- Moore, B. C. J., Alcántara, J. I., and Dau, T (1998). “Masking patterns for sinusoidal and narrow-band noise maskers.” In: *The Journal of the Acoustical Society of America* 104.2 Pt 1, pp. 1023–38.
- Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (2000). “A test for the diagnosis of dead regions in the cochlea”. In: *Br. J. Audiol.* 34, pp. 205–224.
- Moore, B. C. J. and Glasberg, B. R. (2011). “The effect of hearing loss on the resolution of partials and fundamental frequency discrimination”. In: *The Journal of the Acoustical Society of America* 130.5, pp. 2891–2901.
- Moore, B. C. J. and Gockel, H. E. (2011). “Resolvability of components in complex tones and implications for theories of pitch perception”. In: *Hearing Research* 276.1-2, pp. 88–97.
- Moore, B. C. J. and Moore, G. A. (2003). “Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects”. In: *Hearing Research* 182.1-2, pp. 153–163.
- Moore, B. C. J., Glasberg, B. R., and Vickers, D. A. (1995). “Simulation of the effects of loudness recruitment on the intelligibility of speech in noise”. In: *British journal of audiology* 29, pp. 131–143.
- Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (1996). “Effect of loudness recruitment on the perception of amplitude modulation”. In: *The Journal of the Acoustical Society of America* 100.1, pp. 481–489.
- Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (1999). “Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear

- active mechanism". In: *The Journal of the Acoustical Society of America* 106.5, pp. 2761–2778.
- Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (2006a). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure." In: *The Journal of the Acoustical Society of America* 119.1, pp. 480–490.
- Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (2006b). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure". In: *Hearing Research* 222.1-2, pp. 16–27.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). "Musicians have enhanced subcortical auditory and audiovisual processing of speech and music." In: *Proceedings of the National Academy of Sciences of the United States of America* 104.40, pp. 15894–15898.
- Naber, M. and Nakayama, K. (2013). "Pupil responses to high-level image content". In: *Journal of vision* 13.2013, pp. 1–8.
- Nelson, D. A., Schroder, A. C., and Wojtczak, M. (2001). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners." In: *The Journal of the Acoustical Society of America* 110.October, pp. 2045–2064.
- Norman-Haignere, S., Kanwisher, N., and McDermott, J. H. (2013). "Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex". In: *Journal of Neuroscience* 33.50, pp. 19451–19469.
- Norman-Haignere, S. and McDermott, J. H. (2016). "Distortion products in auditory fMRI research: Measurements and solutions". In: *NeuroImage* 129, pp. 401–413.
- Ohm, G. (1843). "On the definition of a tone with the associated theory of the siren and similar sound producing devices". In: *Poggendorf's Annalen der Physik und Chemie* 59.497ff.
- Oxenham, A. J. and Plack, C. J. (1997). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing." In: *The Journal of the Acoustical Society of America* 101.6, pp. 3666–3675.

- Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (2011). "Pitch perception beyond the traditional existence region of pitch". In: *Proceedings of the National Academy of Sciences* 108.18, pp. 7629–7634.
- Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G. (2003). "Informational masking and musical training." In: *The Journal of the Acoustical Society of America* 114.3, pp. 1543–1549.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). "Increased auditory cortical representation in musicians." In: *Nature* 392.6678, pp. 811–814.
- Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). "Timbre-specific enhancement of auditory cortical representations in musicians." In: *Neuroreport* 12.1, pp. 169–174.
- Pantev, C. and Herholz, S. C. (2011). "Plasticity of the human auditory cortex related to musical training". In: *Neuroscience and Biobehavioral Reviews* 35.10, pp. 2140–2154.
- Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., and Schulz, M. (2003). "Music and Learning-Induced Cortical Plasticity". In: *Annals of the New York Academy of Sciences* 999, pp. 438–450.
- Parbery-Clark, A., Tierney, A., Strait, D. L., and Kraus, N. (2012). "Musicians have fine-tuned neural distinction of speech syllables". In: *Neuroscience* 219, pp. 111–119.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). "Musical experience limits the degradative effects of background noise on the neural processing of sound." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.45, pp. 14100–7.
- Patterson, R. D. (1976). "Auditory Filter Shapes Derived with Noise Stimuli". In: *The Journal of the Acoustical Society of America* 59.3, pp. 640–654.
- Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). "A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging". In: *Journal of Neuroscience* 24.30, pp. 6810–6815.

- Piquado, T., Isaacowitz, D., and Wingfield, A. (2010). "Pupillometry as a measure of cognitive effort in younger and older adults". In: *Psychophysiology* 47.3, pp. 560–569.
- Plack, C. (2005). *The sense of hearing*. New Jersey: Erlbaum.
- Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N., eds. (2005). *Pitch: Neural Coding and Perception*. New York: Springer.
- Plack, C. J., Barker, D., and Hall, D. A. (2014). "Pitch coding and pitch processing in the human brain". In: *Hearing Research* 307. August, pp. 53–64.
- Pook, G. (1973). "Information processing versus pupil diameter". In: *Percept Mot Skills* 37, pp. 1000–1002.
- Porter, G., Troscianko, T., and Gilchrist, I. D. (2007). "Effort during visual search and counting: Insights from pupillometry". In: *The Quarterly Journal of Experimental Psychology* 60.2, pp. 211–229.
- Puschmann, S., Uppenkamp, S., Kollmeier, B., and Thiel, C. M. (2010). "Dichotic pitch activates pitch processing centre in Heschl's gyrus." In: *NeuroImage* 49.2, pp. 1641–9.
- Puschmann, S., Özyurt, J., Uppenkamp, S., and Thiel, C. M. (2013). "Pitch-induced responses in the right auditory cortex correlate with musical ability in normal listeners." In: *Neuroreport* 24, pp. 841–5.
- Ruggero, M. A. (1992). "Responses to sound of the basilar membrane of the mammalian of cochlea". In: *Current Opinion in Neurobiology* 2, pp. 449–456.
- Rutherford, E. (1886). "A new theory of hearing". In: *J Anat Physiol* 21, pp. 166–168.
- Santurette, S. (2011). "Neural coding and perception of pitch in the normal and impaired human auditory system". PhD thesis. Lyngby: Technical University of Denmark.
- Scheidt, R. E., Kale, S., and Heinz, M. G. (2010). "Noise-induced hearing loss alters the temporal dynamics of auditory-nerve responses". In: *Hearing Research* 269.1-2, pp. 23–33.
- Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A., and Winner, E. (2009). "Training-induced neuroplasticity in young children". In: *Annals of the New York Academy of Sciences* 1169, pp. 205–208.

- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). "Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians." In: *Nature neuroscience* 5.7, pp. 688–694.
- Schouten, J., Ritsma, R., and Lopez Cardozo, B. (1962). "Pitch of the Residue". In: 294.1940, pp. 1418–1424.
- Schwarz, G. (1978). "Estimating the dimension of a model". In: *The annals of statistics* 6.2, pp. 461–464.
- Sek, A. and Moore, B. C. J. (1994). "The critical modulation frequency and its relationship to auditory filtering at low frequencies." In: *The Journal of the Acoustical Society of America* 95.5 Pt 1, pp. 2606–2615.
- Shackleton, T. M. and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination." In: *The Journal of the Acoustical Society of America* 95.6, pp. 3529–3540.
- Shamma, S and Klein, D (2000). "The case of the missing pitch templates: how harmonic templates emerge in the early auditory system." In: *The Journal of the Acoustical Society of America* 107.5 Pt 1, pp. 2631–2644.
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., and Roberts, N. (2002). "Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians". In: *NeuroImage* 17.3, pp. 1613–1622.
- Soderquist, D. R. (1970). "Frequency analysis and the critical band". In: *Psychonomic Science* 21.2, pp. 117–119.
- Spiegel, M. F. and Watson, C. S. (1984). "Performance on frequency-discrimination tasks by musicians and nonmusicians". In: *The Journal of the Acoustical Society of America* 76.December, p. 1690.
- Sundberg, J. (1977). "The acoustics of the singing voice." In: *Scientific American* 236.3, pp. 104–116.
- Terhardt, E. (1974). "Pitch, consonance and harmony". In: *Journal of the Acoustical Society of America* 55, pp. 1061–1069.
- Wendt, D., Dau, T., and Hjortkjær, J. (2016). "Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension". In: *Frontiers in Psychology* 7.March, pp. 1–12.

- Wightman, F. (1973). "The pattern-transformation model of pitch". In: *Journal of the Acoustical Society of America* 54, pp. 407–416.
- Wolfe, J., Garnier, M., and Smith, J. (2009). "Vocal tract resonances in speech, singing, and playing musical instruments." In: *HFSP journal* 3.1, pp. 6–23.
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). "Musical experience shapes human brainstem encoding of linguistic pitch patterns." In: *Nature neuroscience* 10.4, pp. 420–422.
- Zatorre, R. J. and Belin, P. (2001). "Spectral and temporal processing in human auditory cortex." In: *Cerebral cortex* 11.10, pp. 946–953.
- Zatorre, R. J. (1988). "Pitch perception of complex tones and human temporal-lobe function". In: *Journal of the Acoustical Society of America* 84.2, pp. 566–572.
- Zatorre, R. J. and Zarate, J. M. (2012). *The Human Auditory Cortex*. Vol. 43, pp. 261–294.
- Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). "Neural mechanisms underlying melodic perception Memory for Pitch". In: *Journal of Neuroscience* 14.4, pp. 1908–1919.
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). "Structure and function of auditory cortex: music and speech". In: *Trends in Cognitive Sciences* 6.1, pp. 37–46.
- Zatorre, R. J., Delhommeau, K., and Zarate, J. M. (2012). "Modulation of auditory cortex response to pitch variation following training with microtonal melodies". In: *Frontiers in Psychology* 3.544, pp. 1–17.
- Zekveld, A. A. and Kramer, S. E. (2014). "Cognitive processing load across a wide range of listening conditions: Insights from pupillometry". In: *Psychophysiology* 51.3, pp. 277–284.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). "Pupil response as an indication of effortful listening: the influence of sentence intelligibility." In: *Ear & Hearing* 31.4, pp. 480–490.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). "Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response." In: *Ear & Hearing* 32.4, pp. 498–510.

A

Experimental Evidence for a Cochlear Source of the Precedence Effect^a

Abstract

The precedence effect (PE) refers to the dominance of directional information carried by a direct sound (lead) over the spatial information contained in its multiple reflections (lags) in sound localization. Although the processes underlying the PE have been largely investigated, the extent to which peripheral versus central auditory processes contribute to this perceptual phenomenon has remained unclear. The present study investigated the contribution of peripheral processing to the PE through a comparison of physiological and psychoacoustical data in the same human listeners. The psychoacoustical experiments, comprising a fusion task, an interaural time difference detection task and a lateralization task, demonstrated a time range from 1 to 4.6-5 ms, in which the PE operated (precedence window). Click-evoked otoacoustic emissions (CEOAEs) were recorded in both ears to investigate the lead-lag interactions at the level of the basilar membrane (BM) in the cochlea. The CEOAE-derived peripheral and monaural lag suppression was largest for ICIs of 1-4 ms. Auditory-evoked brainstem responses (ABRs) were used to investigate monaural and binaural lag suppression at the brainstem level. The responses to monaural stimulation reflected the peripheral lag suppression observed in the CEOAE results, while the binaural brainstem responses did not show

^a This appendix is based on Bianchi et al. (2013), *J. Assoc. Res. Otolaryngol.* 14, pp. 767-779.

any substantial contribution of binaural processes to monaural lag suppression. The results demonstrated that the lag suppression occurring at the BM in a time range from 1 to 4ms, as indicated by the suppression of the lag-CEOAE, was the source of the reduction in the lag-ABRs and a possible peripheral contributor to the PE for click stimuli.

A.1 Introduction

In an enclosed environment, the signal generated from a sound source reaches the listener both through a direct path and from multiple reflections off the room's surfaces. Although the listener receives reflections from different locations, the auditory system is generally able to localize the sound source rather accurately by suppressing the directional cues carried by the numerous reflections. The perceptual phenomenon of dominance of the directional information contained in the first arriving sound is known as the precedence effect (PE) (Wallach et al. 1949; Zurek 1987). This natural situation of a direct sound followed by multiple reflections can be simplified by considering a direct sound with a single reflection. The direct sound (lead) and its reflection (lag) can be reproduced in the free field by two loudspeakers at different locations, driven with identical click stimuli with a delay between the onsets (lead-lag delay or inter-click interval (ICI)). The perception of the lead-lag pair depends on the ICI and varies both in the number of perceived stimuli and in their perceived location. Although this variation is gradual and stimulus dependent, some approximate ranges of perception can be defined: a summing window, a precedence window, and an echo window (Fitzpatrick et al. 1999; Litovsky et al. 1999). The summing window is defined by an ICI range between 0 and 1 ms (e.g., Litovsky et al. 1999), where the lead and the lag are perceptually fused in one single image and contribute both to the perceived localization of the fused event. The precedence window is defined by an ICI range from 1 ms up to the echo threshold (Fitzpatrick et al. 1999; Litovsky et al. 1999). Here, the percept is a fused event localized at the lead location. For this time range, the directional cues contained in the lag are weighted less heavily than those of

the lead (Wallach et al. 1949; Litovsky et al. 1999). The echo window refers to the ICI range above the echo threshold, where the lead and the lag are audible as two separated sound images, each perceived at its own location (Blauert 1997). The echo threshold estimates the ICI at which the fused auditory event perceptually splits into two sound images. For clicks, the echo threshold occurs at ICIs of 2-10 ms (Freyman et al. 1991; Yang and Grantham 1997b; Litovsky et al. 1999), and studies using headphones generally observe smaller values (2-4 ms) than those using loudspeakers (Fitzpatrick et al. 1999; Litovsky et al. 1999). Although the PE has been intensively studied over the last two decades (Lindemann 1986; Divenyi and Blauert 1987; Freyman et al. 1991; Fitzpatrick et al. 1995; Litovsky and Yin 1998; Fitzpatrick et al. 1999; Liebenthal and Pratt 1999; Hartung and Trahiotis 2001; Damaschke et al. 2005; Xia and Shinn-Cunningham 2011), the debate whether the lag-suppression mechanism results from peripheral or central processes has remained unresolved. Previous studies have suggested the existence of monaural and peripheral mechanisms responsible for a reduction in the sensitivity to the spatial cues contained in the lagging stimulus (Tollin 1998; Tollin and Henning 1998, 1999; Hartung and Trahiotis 2001; Wolf et al. 2010; Xia and Shinn-Cunningham 2011). However, these studies either consisted of solely psychoacoustical experiments (Tollin and Henning 1998, 1999), a test of computational models against psychoacoustical results (Tollin 1998; Hartung and Trahiotis 2001; Xia and Shinn-Cunningham 2011) or physiological findings in animals (Wolf et al. 2010). Monaural neural correlates of lag suppression were also reported by Wickesberg and Oertel (1990), Fitzpatrick et al. (1995), Parham et al. (1996), Fitzpatrick et al. (1999), and Tollin et al. (2004). The current study investigated contributions to the PE at different stages along the auditory pathway, whereby comparisons between psychoacoustical and physiological data were analyzed in the same human listeners. Three psychoacoustical experiments, a fusion task, an interaural time difference (ITD) detection task, and a lateralization task were performed to investigate the perceptual phenomena related to the PE. Furthermore, noninvasive physiological methods, click-evoked otoacoustic emissions (CEOAEs), and auditory-evoked brainstem responses (ABRs), were used to systematically examine the effect of the leading click on the lagging click at cochlear and brainstem levels

and to experimentally test the hypothesis of a peripheral source of the PE.

A.2 Methods

Six normal-hearing subjects (three females and three males), aged from 24 to 34, participated in the experiments. All had audibility thresholds of less than 20 dB hearing level at the frequencies in a standard audiogram. The experiments took place in a doublewalled soundproof booth that was electrically shielded for the CEOAE and ABR experiments. All signals were generated digitally in MATLAB at a sampling rate of 48 kHz and consisted of 83 μ s clicks.

A.2.1 Psychoacoustical experiments

The psychoacoustical experiments investigated two perceptual phenomena that characterize the perception of the lead-lag pair in the precedence window (Litovsky et al. 1999): fusion, which refers to the perception of one single, fused auditory event and lag-discrimination suppression, which refers to the difficulty of the listener to discriminate directional information contained in the lag. The stimuli, consisting of lead-lag click pairs of the type presented in Figure A.1a, were presented over headphones (Sennheiser HD580) using a D/A converter (type RME DIGI96/8 PAD). The lead-lag pairs were presented at 75 dB peak equivalent sound pressure level (peSPL) and had ICIs of 1, 2, 3, 4, 5, and 8 ms. Two stimulus conditions were considered: a reference condition (ITD=0, lead and lag perceived at the center of the head; Fig. A.1a left) and a deviant condition (lag-ITD>0, lag lateralized towards the left; Fig. A.1a right).

Fusion test

An adaptive one-interval, two-alternative forced-choice (2 AFC) procedure was adopted to determine the echo threshold, i.e., the ICI for which the deviant was perceived as two separate clicks. Each presentation consisted of a deviant with a lag-ITD of 300 μ s, for which the ICI was varied between 1 and 7 ms. The test was carried

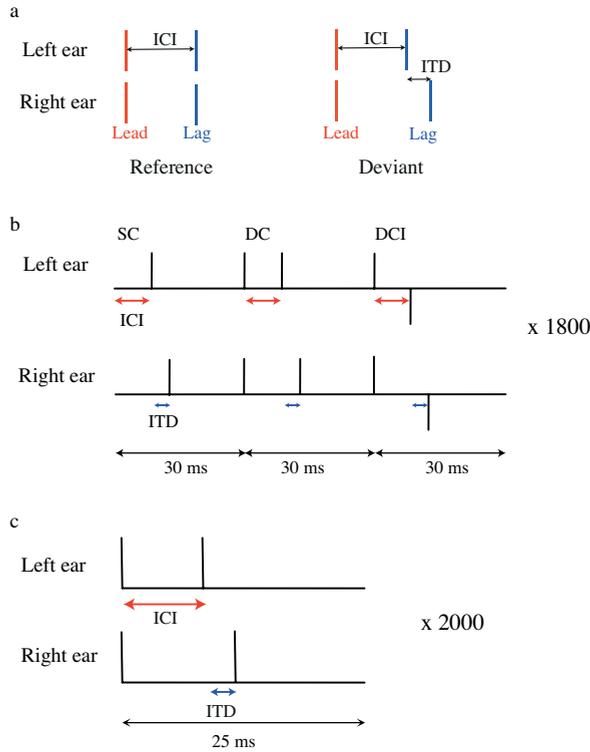


Figure A.1: a) Schematic stimulus configurations used in the behavioral experiments: reference and deviant. The reference configuration consists of two diotic click pairs ($ITD=0$), delayed by an inter-click interval (ICI). In the deviant configuration, the lead is represented as a diotic click pair ($lead-ITD=0$) and the lag as a dichotic click pair ($lag-ITD>0$). b) Interleaved stimulus presentation used in the CEOAE experiment. Three configurations (SC single click; DC double click; DCI double-click inverted) were repeated 1800 times within a sequence for each ICI condition and for an ITD of $300\ \mu s$. c) Stimulus presentation for the ABR experiment. A deviant configuration was repeated 2000 times, for each ICI condition and for an ITD of $300\ \mu s$. The ABRs were recorded by using four electrodes: Fz (ground, positioned at the forehead), Cz (reference, positioned at the vertex), and M1 and M2 (left and right mastoids).

out both for monaural and binaural stimulation to investigate the contribution of binaural processing to fusion. In both tests, the subjects' task was to specify whether they perceived a single click (SC; fused image) or two separated clicks (lead and lag). The subjects were instructed to press the two-click response only when they could hear two auditory events clearly separated in time (monaural test) or in space (binaural test). The starting value of the ICI was 1 ms, which was increased after each single-click response and was decreased after two consecutive two-click responses. The initial step size was 1 ms and reduced after a lower reversal to 0.5 and 0.3 ms as the threshold was approached. The echo threshold was obtained after six reversals and corresponded to the 70.7 % point on the psychometric function. Thresholds were obtained as the average of three repeated measurements.

ITD-detection test

This test investigated lag-discrimination suppression by studying lag-ITD detection as a function of ICI. Seven sequences containing references and deviants were presented, one for each of the following ICIs: 0, 1, 2, 3, 4, 5, and 8 ms. Within each sequence, the ICI was constant and the deviants were randomly presented among the references, allowing a minimum of three references between the presentation of two deviants (Damaschke et al. 2005). The deviants contained ITDs ranging from 150 to 900 μ s with a step size of 150 μ s. Each ITD was repeated three times within the same sequence for a total of 18 deviants per sequence (six lag-ITDs repeated three times). The interval between the onset of one lead-lag pair and the onset of the following pair was 1 s. The subjects' task was to hit a button on the keyboard whenever a noncentered click pair (i.e., a deviant) was detected among the centered references. The response was considered correct when the button was pressed within 1 s after the presentation of the deviant. False alarms were accounted for by calculating the ratio between the number of correct hits and the total hits for each sequence. Subjects were asked to repeat those sequences where the ratio was below 70 %. The ITD-detection threshold was calculated as the lag-ITD that corresponded to 67 % correct performance, i.e., when the lag-ITD was correctly detected at least two times out of three for each sequence.

Lateralization test

The stimulus presentation consisted of one interval containing two lead-lag pairs: a reference followed by a deviant. The reference and deviant had the same ICI, with values among: 0, 1, 2, 3, 4, 5, or 8 ms. The deviant contained a lag-ITD in the right channel, which was randomly varied among: 0, 150, 300, 450, 600, 750, 900, and 1,000 μ s. Each ITD was repeated three times for each ICI. After the presentation of each reference-deviant pair, subjects were asked to press one of the six response buttons ((1) left, (2) center, (3) center and center, (4) center and right, (5) center and left, and (6) center and left and right) according to the perceived lateralization of the deviant with respect to the reference. The six buttons were designed to take all possible percepts of the deviant into account, both when fusion occurred and when fusion was no longer present. In the case of a fused percept, a SC was perceived, either to the left (when the ITD was detected) or at the center. Otherwise, lead and lag were perceived as two separate clicks, where the lead was always perceived as centered, and the lag was perceived either at the center, left, or right, or as two clicks to the left and to the right. Although the lag ITD was leading to the left ear, the percepts of the lag either to the right, or to the left and right, were included to account for the possibility of different monaural suppressions of the lagging clicks in the left and right ear (e.g., for large ITDs). The lateralization threshold was calculated for each ICI as the minimum ITD producing at least two times out of three (67 %) a noncentered percept of the deviant.

CEOAE recordings

The stimuli were sent via the open source software pwavplay to the soundcard (RME FireFace 800 A/D-D/A converter, RME Intelligent Audio Solutions, Germany). The clicks were calibrated at a level of 65 dB peSPL in a BK-2012 ear-canal coupler (Bruel & Kjaer Sound & Vibration Measurement A/S, Denmark), attached to a BK-4157 artificial ear. After insertion of the recording probe in the ear canal, in situ calibration was performed using a TDT-PA5 programmable attenuator (Tucker-Davis Technologies, Alachua, FL) to ensure that the levels of the clicks in the ear canal were equal in each ear. The stimuli were presented to the left and right ear

of the test subjects via two ER-2 earphones (Etymotic Research, Inc., Elk Grove Village, IL). Recordings were performed using two ER-10B+ low-noise microphones and were bandpass filtered between 0.6 and 5 kHz (analog Rockland 852 HI/LO filter). Click pair stimuli were designed for seven different ICIs (0, 1, 2, 3, 4, 5, and 8 ms) and a lag-ITD of 300 μ s. The response recorded to the double-click stimulus consists of a CEOAE to the lead click, a CEOAE to the lag click, and a nonlinear component that depends on the ICI (Verhulst et al. 2011a). Kemp and Chum (1980a) developed a technique to remove the CEOAE component from the leading click while keeping the CEOAE component to the lagging click and the nonlinear component due to the ICI. This technique, as adapted by Kapadia and Lutman (2000b), was used here to calculate the derived suppressed (DS) response of the lagging click. Figure A.1b illustrates this interleaved procedure adopted for stimulus presentation (Verhulst et al. 2011a). For each ICI and ITD condition, 1,800 repetitions of the following three stimuli were presented: SC, double click (DC; two condensation clicks), and double-click inverted (DCI; one condensation and one rarefaction click). The unsuppressed response (US) corresponded to the SC recordings. The DS response was obtained by subtracting the DCI response from the DC response and by halving the result. The DS response thus consisted of the CEOAE component due to the lagging click and the nonlinear component due to the ICI. The lag suppression was calculated as the root-mean-square (rms) level difference between DS and US responses in a time frame of 6-18 ms after click onset. Both monaural and binaural stimulations were tested. As no difference in lag suppression level was found between the two stimulations, it was decided to present the stimuli binaurally to extract monaural CEOAE lag suppression.

ABR recordings

The electrodes were placed according to the 10-10 system (American Clinical Neurophysiology Society), using a tight-fitting elastic cap that holds the electrodes in position (Picton 2011). Four electrodes were used: Cz (at the vertex, halfway between nasion and inion), Fz (at the forehead at three tenths of nasion- inion distance), M1 (left mastoid), and M2 (right mastoid). The electrode Cz was used

as a reference and the electrode Fz as ground. Low impedances (below 2 k Ω) were achieved by carefully degreasing the test subject's scalp with alcohol and an abrasive electrolyte gel. The stimuli were played back and sent to the soundcard (RME FireFace 800 D/A converter, RME Intelligent Audio Solutions, Germany). The clicks were calibrated at a level of 75 dB peSPL in a BK-2012 ear-canal coupler, attached to a BK-4157 artificial-ear calibrator. The stimuli were presented to the left and right ear of the test subjects via two ER-2 earphones (Etymotic Research, Inc., Elk Grove Village, IL). The electrodes were connected to an EEG amplifier (Synamps 5803), responsible for the amplification and A/D conversion of the recorded potentials. The output of the amplifier was connected to the recording PC where the EEG-data were postprocessed. The average, variance, and covariance of the evoked responses were calculated, and the resulting waveform was bandpass filtered with a FIR filter with cut-off frequencies of 200 and 1,500 Hz. Deviants were presented for seven different ICIs (0, 1, 2, 3, 4, 5, and 8 ms) and a lag-ITD of 300 μ s. For each ICI and ITD condition, the 25-ms-long epoch containing the deviant stimulus was presented 2,000 times (Fig. A.1c). In the data analysis, the wave V amplitude peaks of the lead were determined as the maximum voltage (absolute value) in a time range of 6.5-7.5 ms after stimulus onset (Damaschke et al. 2005). The wave V amplitude peaks of the lag were determined with a similar procedure, in a time range shifted in latency according to the ICI and the ITD.

A.2.2 Statistical analysis

CEOAE

The data obtained for the DS and US conditions were divided into five blocks of 360 averages each. Mean and rms level were calculated for each block and suppression was calculated for the 25 combinations of level difference between the DS and US conditions. The standard deviation (SD) was calculated over the 25 values of suppression (Verhulst et al. 2011a).

ABR

SDs of the ABR recordings were calculated as the square root of the time-averaged variances. Normal distributions were built from the mean and SD of the wave-Vs of lead and lag. A normal distribution of lag-wave V suppression and its SD were obtained by random sampling from the distributions of the lead and lag wave-Vs.

Confidence interval and significance testing

For each subject, a statistical analysis was carried out to investigate whether the CEOAE-derived and the ABR-derived lag suppression was significantly different below and above the individual echo thresholds (Table A.1). For each subject, mean values of lag suppression below and above the echo threshold were calculated from all data points below and above the threshold, respectively. SDs of the mean lag suppression below and above the threshold were obtained by taking the square root of the summed variances, divided by the number of data points (Bienaymé formula). Two normal distributions for data below and above the echo threshold were built from the calculated mean and SDs, and 10,000 random samples were then drawn from each distribution. These two sets of random samples were subtracted to obtain an estimate of the difference distribution of lag suppression below vs. above the echo threshold, and 95 % confidence intervals (CIs) were calculated for these difference distributions. As the sample size of CEOAE and ABR recordings differed, a conservative approach was adopted such that the CIs were defined as the mean of each difference distributions ± 1.96 SD. Significance testing was carried out by controlling whether the CIs contained zero. CIs that did not contain zero (asterisks in Table A.1) indicated that lag suppression was significantly larger above than below the echo threshold. The indicated p values were calculated using the z statistic as $p = e^{-0.717z - 0.416z^2}$ (Altman and Bland 2011).

Table A.1: Mean lag-suppression and standard deviation [dB], calculated for each subject, for ICIs below and above the individual echo-thresholds. Results are shown for OAE and ABR measurements (monaural and binaural stimulation) and a fixed ITD of $300 \mu s$. The third column shows the 95% confidence intervals (CI; lower and upper limits) of the difference distributions of lag-suppression below vs. above the echo-threshold.

Exp.	Subj.	Lag-suppression (dB)		95% CI (dB)		p-value
		Below echo-thr	Above echo-thr	lower limit	upper limit	
OAE mono L	KE	4.9 ± 0.2	1.7 ± 0.2	2.6	3.7	<0.0001***
	AL	4.7 ± 0.1	1.0 ± 0.1	3.4	4.1	<0.0001***
	EC	3.8 ± 0.2	1.4 ± 0.3	1.6	3.1	<0.0001***
	FB	3.9 ± 0.1	1.0 ± 0.1	2.5	3.2	<0.0001***
	NL	3.3 ± 0.3	1.7 ± 0.4	0.7	2.6	0.001**
	SV	4.5 ± 0.1	0.3 ± 0.2	3.7	4.6	<0.0001***
OAE mono R	KE	6.6 ± 0.1	1.4 ± 0.2	4.7	5.6	<0.0001***
	AL	4.0 ± 0.1	1.2 ± 0.2	2.3	3.3	<0.0001***
	EC	4.6 ± 0.2	2.5 ± 0.2	1.7	2.6	<0.0001***
	FB	4.3 ± 0.1	1.6 ± 0.1	2.4	3.2	<0.0001***
	NL	4.2 ± 0.7	2.0 ± 0.3	0.7	3.6	0.004**
	SV	3.8 ± 0.1	0.8 ± 0.2	2.6	3.4	<0.0001***
ABR mono L	KE	4.7 ± 1.9	-0.3 ± 0.5	1.2	9.0	0.0106*
	AL	4.7 ± 2.0	1.8 ± 1.2	-1.7	7.6	0.2134
	EC	3.3 ± 1.8	0.7 ± 0.7	-1.2	6.4	0.174
	FB	4.3 ± 2.5	0.3 ± 0.6	-1.1	8.9	0.1259
	NL	2.1 ± 2.1	-0.3 ± 1.0	-2.1	6.9	0.2994
	SV	3.4 ± 1.3	1.9 ± 0.9	-1.7	4.7	0.3766
ABR mono R	KE	2.2 ± 0.6	-0.2 ± 0.3	1.1	3.8	0.0006***
	AL	3.0 ± 0.6	0.8 ± 0.5	0.6	3.7	0.0071**
	EC	3.7 ± 2.0	2.1 ± 0.7	-2.5	5.6	0.4673
	FB	2.9 ± 1.8	2.0 ± 2.5	-5.3	6.9	0.808
	NL	3.4 ± 3.4	0.4 ± 1.0	-3.9	10.1	0.3949
	SV	2.7 ± 0.6	0.7 ± 0.6	0.3	3.7	0.0192*
ABR binaural	KE	3.2 ± 0.5	0.6 ± 0.2	1.5	3.5	<0.0001***
	AL	4.6 ± 1.4	1.1 ± 0.3	0.8	6.3	0.0124*
	EC	6.2 ± 1.8	0.8 ± 0.3	1.8	9.1	0.0036**
	FB	2.0 ± 0.8	0.6 ± 0.5	-0.4	3.3	0.1162
	NL	3.9 ± 1.8	1.2 ± 0.7	-1.1	6.4	0.167
	SV	4.4 ± 1.4	1.6 ± 0.2	-0.1	5.7	0.057

Asterisks denote CI significantly larger than zero: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

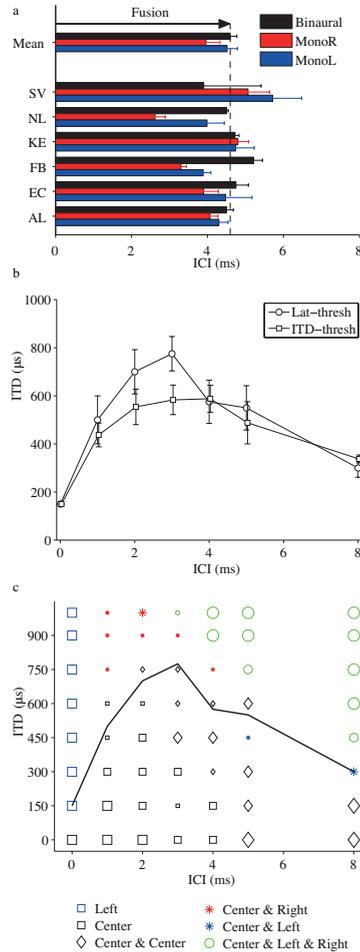


Figure A.2: Psychoacoustical results. a) Individual and mean results of the fusion test for binaural (black bars) and monaural stimulation (monaural right, red bars; monaural left, blue bars) by deviants with a lag-ITD of $300 \mu\text{s}$. b) Mean behavioral thresholds obtained from the lateralization test (circles) and ITD-detection test (squares). The error bars represent the standard error of the mean. c) Lateralizations reported the most by the six subjects over three repetitions of the lateralization test (symbols) and mean lateralization threshold (black curve). The different markers represent the six response buttons (left, center, center and center, center and left, center and right, and center and left and right). The size of the symbols indicates at what percentage the lateralization was reported over 18 responses (six subjects, three repetitions): small symbols, below 50 %; medium symbols, between 50 and 70 %; large symbols, above 70 %.

A.3 Results

A.3.1 Psychoacoustical experiments

The individual and mean results of the fusion test are presented in Figure A.2a. The figure shows the ICIs for which fusion occurred, both for monaural (monaural left, blue bar; monaural right, red bar) and binaural stimulation (black bar). The breakdown of fusion corresponds to the echo threshold. The mean results show similar echo thresholds for binaural stimulation (4.6 ms) and monaural left stimulation (4.5 ms). For monaural right stimulation, a lower value of 4 ms was observed, due to the additional delay of 300 μ s introduced by the ITD. The similar values for the echo thresholds obtained in the monaural and binaural conditions suggest a fusion mechanism that does not depend on binaural processes. This is consistent with other studies where similar echo thresholds were found in the absence and presence of binaural cues (Rakerd et al. 1997) and for subjects with monaural deafness and normal-hearing subjects (Litovsky et al. 1997). Figure A.2b presents the mean ITD-detection thresholds (squares) and lateralization thresholds (circles). The ITD-detection threshold, i.e., the minimum lag-ITD to obtain a noncentered percept of the deviant, increased up to 590 μ s for ICIs between 0 and 4 ms, and then decreased again for ICIs above 4 ms. Large threshold values indicated strong lag-discrimination suppression. For an ICI of 0 ms, no lag-discrimination suppression occurred (i.e., lead and lag had the same weight in lateralization) and all subjects could detect the deviants at the shortest ITD presented (150 μ s). For an ICI of 8 ms, the ITD threshold was 340 μ s, which was significantly higher than the baseline threshold for an ICI of 0 ms ($p < 0.05$, two-sample right-tailed t test) and not significantly lower than the threshold at 5 ms ($p = 0.074$, two-sample right-tailed t test), indicating that lag-discrimination suppression was still present for a lead-lag delay of 8 ms (and ITDs below the threshold). The ITD-threshold obtained here showed an ICI range over which lag-discrimination suppression occurred that is in agreement with previous studies (Zurek 1980; Damaschke et al. 2005). The lateralization test refined the ITD-detection test by specifying the lateralization of a lead-lag pair as a function of the ICI. The difference from the

previous test was that the task in this experiment was not only to detect the ITD contained in the lead-lag pair, but also to specify the perceived lateralization of the lead-lag pair. For each subject, the threshold was calculated as the minimum ITD producing at least two (out of three) noncentered percepts of the deviant. Figure A.2b shows the mean lateralization threshold (circles), where the error bars indicate the standard error of the mean. The lateralization threshold curve presented similar values as the ITD-detection threshold function for all ICIs except at 3 ms, where the lateralization threshold was significantly larger than the detection threshold ($p=0.029$, Wilcoxon rank sum test). Largest thresholds were obtained for ICIs of 2- 3 ms. For longer ICIs, the threshold curve decreased again, until reaching $300 \mu\text{s}$ for an ICI of 8 ms. Although not at baseline level ($150 \mu\text{s}$), this value was significantly lower than the threshold at 5 ms ($p=0.021$, two-sample right-tailed t test). In Figure A.2c, the mean lateralization threshold (black curve) is represented together with the lateralizations that were reported the most by the six subjects. The different symbols represent the different response buttons, whereas the size of the symbols shows at what percentage the lateralization was reported over 18 responses (six subjects and three repetitions). Small symbols indicate the lateralizations that were reported less than nine times (i.e., below 50 and large symbols represent reported lateralizations corresponding to between 50 and 70 respectively. The black symbols indicate perception of the lead-lag pair at the lead location, i.e., when lag-discrimination suppression occurred. Colored symbols show the release from lag-discrimination suppression. Fused percepts are indicated by the squared symbols. For an ICI of 0 ms, the blue squares show that lead and lag had the same weight in lateralization (i.e., summing location), as subjects reported to hear a SC towards the left more than 70 between 1 and 4 ms, lag-ITDs below $600 \mu\text{s}$ show a strong lag-discrimination suppression (black symbols), whereas ITDs above $600 \mu\text{s}$ indicate a release from lag-discrimination suppression (colored symbols), even though difficulties were reported in consistently lateralizing the lag (small symbols). For ICIs above 4 ms, the results for all ITDs indicated that lead and lag were no longer perceived as fused. Despite the breakdown of fusion, lag-discrimination suppression was still observed for ICIs of 5 and 8 ms at short ITDs (black diamonds). For large ITDs, the subjects reported to perceive a diffuse

sound image inside the head (green circles). In summary, the results from the three perceptual experiments estimated fusion to occur within an ICI range up to 4.6 ms, and lag-discrimination suppression to last for longer ICIs (at least up to 8 ms).

A.3.2 CEOAE

When the auditory system is stimulated by a click, the forward travelling wave created along the basilar membrane (BM) can be reflected by preexisting random BM impedance irregularities (Shera and Guinan 1999; Zweig and Shera 1995). These irregularities are inherent to a healthy cochlea and may reflect small cell-to-cell differences in outer-hair cell amplification and alignment, which can be thought of as placefixed BM impedance irregularities. Through a mechanism of coherent reflection, the BM irregularities are assumed to give rise to a backwards traveling wave that can be recorded in the ear canal as a CEOAE (Zweig and Shera 1995). CEOAEs contain information about the BM processing at the cochlear regions where the emission was generated (Moleti et al. 2008; Shera et al. 2002). When the cochlea is stimulated with lead-lag pairs, both the lead and lag elicit a CEOAE. It has been shown that, when preceded by the lead, the CEOAE elicited by the lag is reduced in amplitude compared with a CEOAE elicited by the lag presented in isolation (Kapadia and Lutman 2000; Verhulst et al. 2011a). This CEOAE amplitude reduction, which depends on the lead-lag delay, presumably reflects attenuation of the BM response to the lagging click, and will be referred to as peripheral lag suppression in the following. Figure A.3a shows the spectra of the recorded CEOAEs for one representative subject KE. The spectrum represented in gray is the US, which is the emission elicited by the lag presented in isolation. The superimposed spectrum (white) is the DS response which represents the derived emission of the lag when preceded by the lead. The difference between US and DS (gray region) indicates peripheral lag suppression for three ICI conditions of 2 (left panel), 4 (middle panel), and 8 ms (right panel). The results show that lag suppression was maximal for an ICI of 2 ms and almost negligible for an ICI of 8 ms. Consistent with previous studies (Verhulst et al. (2013); Verhulst et al. 2011b), the figure also shows that the release of lag suppression first occurred at the highest frequencies

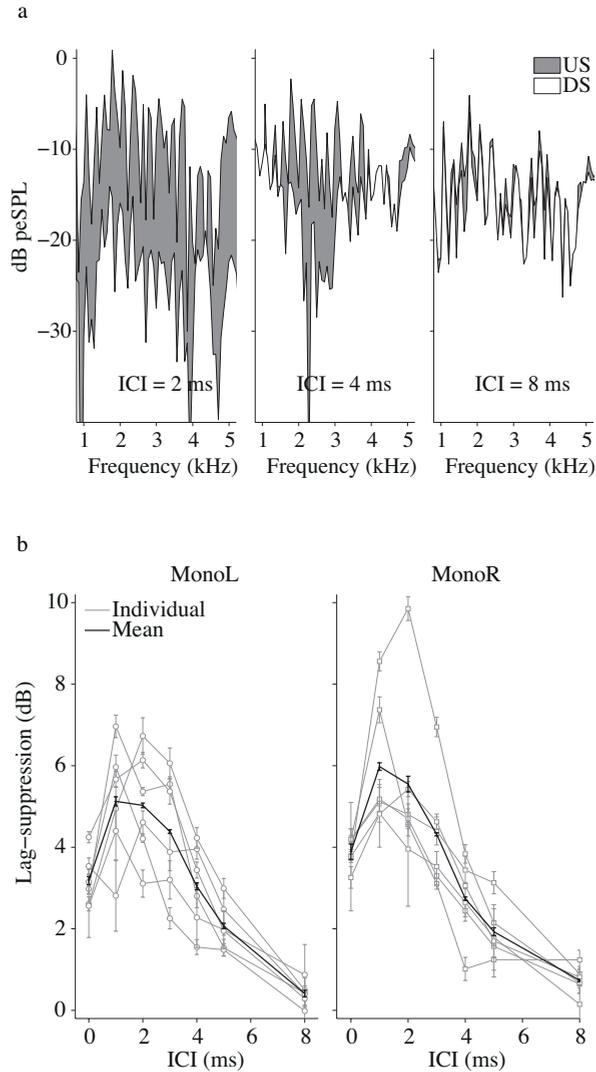


Figure A.3: CEOAE results. a) Spectra of the recorded CEOAEs for the single-click condition, i.e., the unsuppressed response (US), and for the derived suppressed response (DS, obtained from $(DC-DCI)/2$ in Fig. A.1b) of the lagging click, for one representative subject KE. The difference between US and DS (the area displayed in gray) represents peripheral lag suppression for ICIs of 2, 4, 8 ms. b) Individual (gray curves) and mean (black curves) results of peripheral lag suppression as a function of the ICI for monaural left and right stimulation. The error bars indicate the standard error of the mean.

(e.g., at 4 kHz for an ICI of 4 ms), and later at lower frequencies (e.g., at 2 kHz for an ICI of 8 ms). This frequency-dependent release of suppression as a function of ICI appears to be related to BM impulse response duration, where higher characteristic frequencies exhibit a shorter time range of impulse response lead-lag interactions. Thus, the peripheral lag suppression obtained from CEOAE recordings appears to reflect mechanical BM impulse response lead-lag interactions. In Figure A.3b, peripheral lag suppression is represented as a function of ICI. The figure shows individual (gray curves) and mean data (black curves) of peripheral lag suppression for monaural left (left panel) and monaural right (right panel) stimulation, for lead-lag pairs with an ITD of 300 μ s. The mean data show a large suppression of the lag (between 3 and 6 dB) for lead-lag delays up to 4 ms. Above an ICI of 4 ms, the mean peripheral lag suppression decreased to 2 dB at 5 ms and 0.5 dB at 8 ms. A statistical analysis was conducted on the null hypothesis that the difference of individual suppression, calculated for ICIs below and above individual echo thresholds, was zero (95 % CI). All test subjects showed peripheral lag suppression that was significantly larger for ICIs below the individual echo threshold than above it (Table A.1).

A.3.3 ABR

ABRs are auditory-evoked potentials that reflect synchronized neural activity generated at the level of the auditory nerve (AN) and the auditory brainstem. Wave V is typically the most prominent peak in the ABR and is considered to reflect activity stemming from the superior olivary complex in the brainstem (Picton 2011). When stimulating with click pairs, both lead and lag typically elicit a wave V. If the lag suppression obtained in the CEOAEs indeed reflects BM lead-lag interactions, it is expected to obtain an analogue response reduction also in the ABR to monaural stimulation (i.e., in the lag-wave V amplitude). Figure A.4a shows the ABR recordings of one representative subject (KE) to binaural stimulation (black curve, left panel) and monaural stimulations (blue and red curves, right panel), for an ITD of 300 μ s. Wave V amplitude peaks are indicated by downward-pointing triangles. The results show that the leading click evoked a wave V that was constant in am-

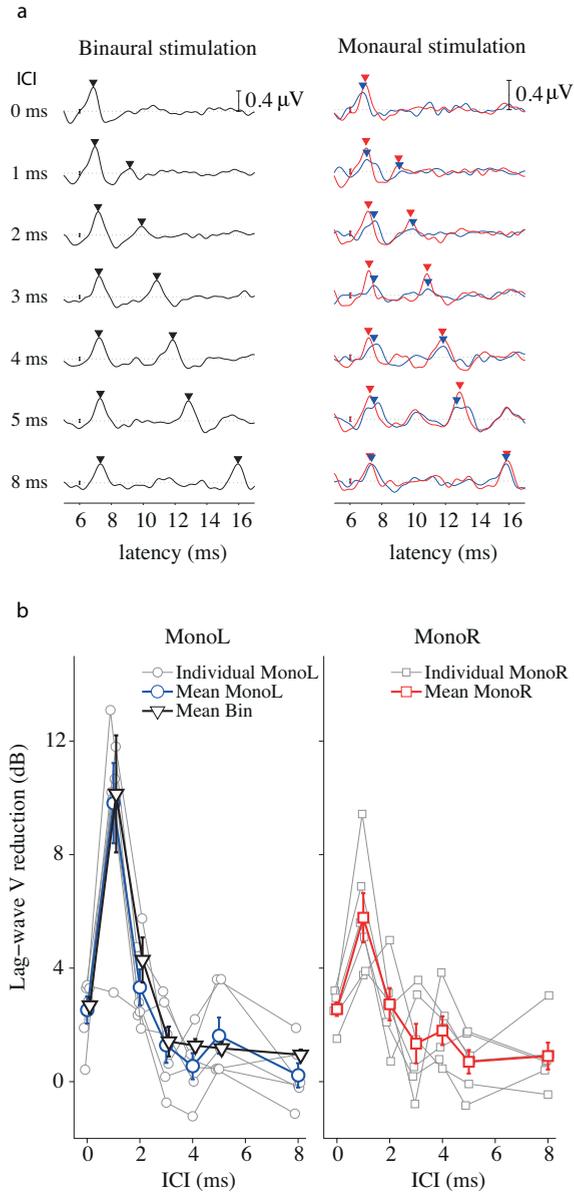


Figure A.4: ABR results. a) ABRs recordings for one representative subject KE, for monaural (right panel, red and blue curves) and binaural (left panel, black curve) stimulation and different ICI conditions. The error bars at a latency of 6ms indicate the time-averaged SD of the recording. The horizontal dashed lines depict the zero voltage reference, and the bar scale at a latency of 16 ms indicates a voltage of $0.4 \mu\text{V}$. b) Individual (gray curves) and mean (black, blue, and red curves) results of lag wave V reduction obtained from ABRs recordings for monaural (gray, blue, and red curves) and binaural stimulation (black curve), as a function of ICI. The error bars indicate the standard error of the mean.

plitude and latency for all ICIs, whereas wave V elicited by the lagging click was initially lower in amplitude for short ICIs and gradually increased in amplitude and latency as ICI increased. Figure A.4b shows individual (gray curves) and mean (black, blue and red curves) lag-wave V reductions as a function of the ICI for monaural left (left panel, blue curve), monaural right (right panel, red curve) and binaural stimulation (left panel, black curve). The mean data show a lag-wave V reduction of up to 10 dB for lead-lag delays of 1 and 2 ms. The reduction obtained for binaural stimulation (black curve, left panel) was not larger than the reduction for monaural left stimulation (blue curve, left panel). A comparison with the behavioral echo thresholds (Table A.1; Fig. A.5) revealed that all subjects showed a lag-wave V reduction that was larger for ICIs below the echo threshold than above it. This result was significant (analysis of 95 % CI of the difference distribution) for three out of six subjects for monaural right stimulation, for one subject for monaural left stimulation, and for three subjects for binaural stimulation (Table A.1).

A.4 Discussion

A.4.1 Effect of frequency range and implications for peripheral processing

Previous studies regarding the auditory processes underlying the PE (Divenyi 1992; Divenyi and Blauert 1987; Dizon and Colburn 2006; Shinn-Cunningham et al. 1995; Tollin and Henning 1999; Wolf et al. 2010; Xia and Shinn-Cunningham 2011) investigated the frequency dependence of localization dominance and lag-discrimination suppression. Two main hypotheses emerged: Divenyi and Blauert (1987) and Blauert and Divenyi (1988) proposed the “spectral overlap” concept, where lag-discrimination suppression was greatest (i.e., ITD thresholds were largest) for a large spectral overlap between the lead and the lag stimuli. Thus, they suggested that discrimination suppression operated within frequency bands (corresponding to peripheral auditory filters). An alternative concept of “localization strength” was proposed by Divenyi (1992) who found that localization dominance decreased with

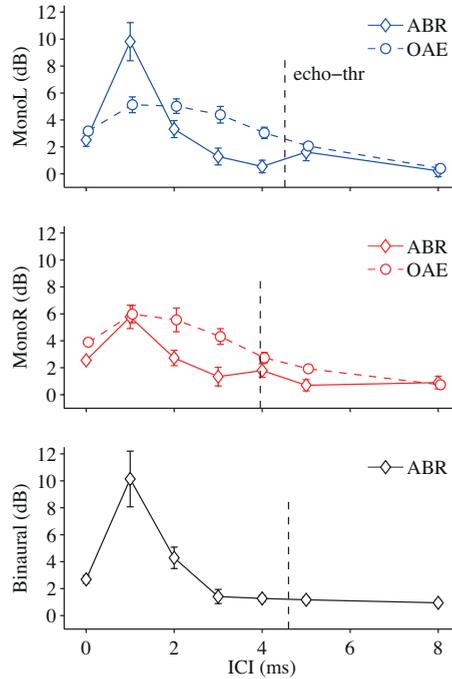


Figure A.5: Comparison of mean lag suppression from OAEs (dashed curves), lag wave V reduction from ABRs (solid curves), and behavioral echo thresholds (vertical dashed lines) for monaural and binaural stimulation. The error bars indicate the standard error of the mean.

decreasing lead center frequency, i.e., a low-frequency lead suppressed the spatial information of a high-frequency lag more strongly than when they were both centered at the same high frequency. This second hypothesis assumed a discrimination suppression mechanism operating across frequency bands. Consistent with the localization strength hypothesis, Shinn-Cunningham et al. (1995) showed that low frequency stimuli dominated over high-frequency stimuli in ITD-detection tasks. Yang and Grantham (1997b) suggested that spectral overlap (i.e., processes operating within frequency bands) and localization strength (i.e., processes across frequency bands) are two independent processes governing discrimination suppression. Other studies investigated the frequency dependence of the PE by using spectrally identical lead and lag stimuli. By varying the center frequency of the

lead-lag pair, these studies investigated within frequencyband effects as a function of frequency. Localization dominance was found to be longer lasting and more pronounced for low frequency lead and lag stimuli than for high frequency stimuli (Lindemann 1986; Tollin and Henning 1999; Dizon and Colburn 2006; Wolf et al. 2010). This frequency-dependent behavior, where localization dominance was demonstrated to decrease with increasing center frequency, strongly supported the contribution of peripheral auditory processing to the PE (Tollin 1998; Hartung and Trahiotis 2001; Wolf et al. 2010; Xia and Shinn-Cunningham 2011). In fact, due to the mechanical properties of the BM, lead and lag exhibit shorter impulse responses and, therefore, shorter interactions when they are both centered at higher frequencies than at lower frequencies. The current study tested this hypothesis experimentally, by measuring CEOAEs to spectrally identical lead and lag stimuli. The results revealed that the CEOAE lag suppression was highly frequency dependent, with longer lasting suppression at low frequencies (Fig. A.3a). By experimentally supporting the previously mentioned studies, these results provide a strong link between BM impulse response duration and within-frequency channel effects reported in psychoacoustical experiments measuring the PE. Although across-frequency processes may also be present, this study shows how within-frequency band lead-lag interactions change over frequency and how this mechanism could affect the perception of a lead-lag pair. The abovementioned studies investigated the frequency dependence of lead dominance and lag-discrimination suppression (i.e., localization tasks), whereas the current study also presented measures of fusion, which does not necessarily involve the extraction of spatial cues. Fusion and discrimination suppression might, to some extent, rely on independent mechanisms, as previously suggested (Yang and Grantham 1997a), and operate in different frequency regions. It has been shown that ITD detection most likely relies on low frequencies (Dizon and Colburn 2006; Tollin and Henning 1999), where the extraction of ITDs is most effective. In contrast, echo thresholds may be dominated by high frequencies, where the lead and lag impulse responses produce shorter interactions on the BM and can, therefore, be separated out for shorter delays than at lower frequencies. The psychoacoustical results of the current study (Fig. A.2) showed slightly different ICI ranges over which fusion and

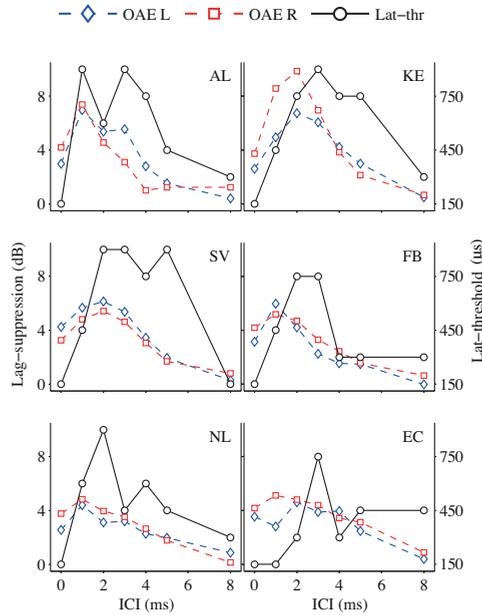


Figure A.6: Individual comparisons of behavioral lateralization thresholds (solid black curves (in microseconds)) and peripheral lag suppression (dashed blue and red curves (in decibels); blue diamonds, monaural left; red squares, monaural right).

lag-discrimination suppression occurred. While fusion broke down at 4.6 ms (Fig. A.2a), lag-discrimination suppression was still strong for an ICI of 5 ms and present for an ICI of 8 ms (for an ITD of 150 μ s, Fig. A.2b, A.2c). The shorter time range over which fusion occurred would, thus, support the hypothesis of dominance of high frequencies for echo threshold determination, where one can extract cues for the number (one or two) of perceived clicks at shorter ICIs than for lateralization.

A.4.2 Effects of peripheral processing on the PE

The CEOAE results (Fig. A.3b) showed that peripheral suppression of the lagging click was maximal for lead-lag delays up to 4 ms, in agreement with previous

studies (Kapadia and Lutman 2000; Verhulst et al. 2011a). For an ICI of 0 ms, the stimulus in the left channel was a SC with double amplitude. Here, no lag suppression occurred and the reduction of 3-4 dB with respect to the single-click condition resulted from the compressive behavior of the CEOAE level curve (Verhulst et al. 2011a). Thus, peripheral lag suppression, defined as the suppressive effect of the lead on the lag, was largest for ICIs between 1 and 4 ms. A comparison of peripheral lag suppression and behavioral monaural echo thresholds (vertical dashed lines) is also presented in Figure A.5. For all test subjects, lag suppression below the echo threshold was significantly larger than that observed above the echo threshold (Table A.1). Figure A.6 shows individual comparisons of peripheral lag suppression (blue and red dashed curves) and behavioral lateralization thresholds (black solid curves). This comparison revealed that large peripheral lag-suppression values were accompanied by higher lateralization thresholds (i.e., when the lagging clicks are monaurally attenuated at the level of the BM, it seems more difficult to lateralize the lag in behavioral tasks). However, while peripheral suppression seems largely responsible for elevating the lateralization thresholds for ICIs of 1- 4 ms, other processes at higher stages may be responsible for raising the thresholds for ICIs of 5 (KE, SV) and 8 ms (thresholds higher than $150 \mu\text{s}$), where OAE and ABR lag suppression was absent. These results provide evidence for a monaural and peripheral component of lag suppression, occurring for lead-lag delays within the precedence window, and suggest a relation between peripheral suppression effects and the perceptual PE. The lag suppression observed in the CEOAEs is of peripheral origin and likely related to the processing at local sites of the BM where the emission was generated. The frequency-dependent release of suppression as a function of ICI (Fig. A.3a) appears to be linked especially to the duration of the local BM impulse response duration, where short ICIs lead to overlapping impulse responses that can cause lag suppression for both low and high frequency cochlear locations, whereas longer ICIs are only able to affect low frequency BM impulse responses. Although there is no invasive study that relates CEOAEs with impulse responses recorded from the BM, a large body of OAE literature provides evidence for spectral components in CEOAEs to reflect local BM processing (Kemp and Chum 1980b; Neely et al. 1988; Zweig and Shera 1995; Shera

and Guinan 1999; Harte et al. 2009). Moreover, cochlear dispersion combined with coherent reflection filtering can explain why the short latencies of the CEOAE waveform contain high frequencies and the longer latencies contain low frequencies (Jedrzejczak et al. 2005; Moleti and Sisto 2008). The above studies support the view that lag suppression observed in CEOAE frequency components can be considered as reflecting complex interactions (both in phase and magnitude) of local BM impulse responses at those cochlear regions where the emission was generated. This view is further supported by two AN studies that performed recordings from single AN fibers to acoustic click pairs (Goblick and Pfeiffer 1969; Parham et al. 1996). While Parham et al. (1996) did not clarify whether the origin of lag suppression arose from adaptation in the AN itself or from cochlear processing that served as an input to the AN, Goblick and Pfeiffer (1969) referred to dynamics in local BM amplification to explain lag suppression. Modeling studies that account for BM as well as higher level processing can provide insight in this matter (Tollin 1998; Hartung and Trahiotis 2001; Xia and Shinn-Cunningham 2011). In the model of Hartung and Trahiotis (2001), two monaural lead-lag stimuli were processed through a left- and right-ear gammatone filterbank (Patterson et al. 1995) and a haircell transduction stage (Meddis 1986) before the outputs were processed by a binaural cross correlation operation. Based on the monaural effects of BM filtering, (inner) hair-cell processing and subsequent binaural processing, the model was shown to qualitatively account for some of the behavioral data associated with the PE (Wallach et al. 1949; Shinn-Cunningham et al. 1995). However, whereas the role of inner-hair-cell (IHC) processing was stressed in the framework of the modeling study, the results from the present study suggest that BM processing, and not IHC/AN processing, might provide the major link between the observed CEOAE-derived lag-suppression data and the behavioral data (in agreement with the model of Tollin 1998). Adaptation effects in the AN and subsequent neural stages may further contribute to the peripheral lag suppression that was shown to affect the perception of the PE in this study. For the click stimuli used in the present study, lag suppression caused by BM impulse response interactions may dominate over AN adaptation effects, which might be stronger for longer duration stimuli.

A.4.3 CEOAEs and monaural ABRs

The mean wave-V amplitude reductions (Fig. A.4b, blue and red curves) obtained from ABR recordings for monaural stimulations were largest in a shorter ICI range (1-2 ms) than the peripheral lag suppression observed in the CEOAEs (Fig. A.3b). Several aspects may account for this difference. First, peripheral lag suppression was measured as an amplitude reduction of the backward travelling wave, which contains information of specific reflection sites along the BM (e.g., Zweig and Shera 1995; Shera et al. 2002). In contrast, the ABR reflects neural activity elicited by the forward travelling wave and, in particular, represents the synchronous activity of neurons across the whole cochlear partition (Dau et al. 2000; Junius and Dau 2005). Even though OAE and ABR results comprise monaural lead-lag interactions, the OAE only contains a subset of frequency components present in the ABR. CEOAEs are, in fact, dominated by frequency components in the 1-2 kHz range where the middle-ear gain is largest (Puria 2003). Moreover, peripheral lag suppression in CEOAEs was observed to be frequency dependent, with longer-lasting suppression at low frequencies than at high frequencies (Verhulst et al. 2011b, Fig. A.3a). Thus, the shorter time range of suppression obtained in the ABR results may be explained by the wider frequency window effective in ABRs versus CEOAEs. Second, ABRs not only reflect outer-hair-cell processing, as in the case of CEOAEs, but also represent effects of IHC processing and neural recovery times in the AN and brainstem.

A.4.4 Contributions of binaural processes

The mean lag-wave V reduction obtained with binaural stimulation (black curve in Fig. A.3b, left panel) was not larger than the one obtained with monaural left stimulation (blue curve). The absence of binaural attenuation at the brainstem is consistent with previous results, which showed correlates of binaural lag suppression only in middle-latency responses but not in early latency responses (Liebenthal and Pratt 1999), and with results showing correlates of binaural lag suppression in the pattern of late auditory-evoked potentials (Damaschke et al. 2005). Although the present study is in agreement with the absence of a binaural contribution to lag suppression at the brainstem level (Damaschke et al. 2005), the conclusion

here differs with respect to the monaural mechanism occurring for stages below the brainstem. While previous studies (Damaschke et al. 2005; Fitzpatrick et al. 1999) concluded that monaural lag-suppression mechanisms occurring for ICIs below 5 ms originate from recovery times in neurons of the AN and brainstem, the present study presents evidence for mechanical BM lead-lag interactions as the main source of lag suppression for ICIs between 1 and 4 ms. When the cue for lateralization is carried by the lag, a mechanism of monaural suppression would account for the raise in the lateralization threshold for short ICIs. This is consistent with results from a recent study (Fisher et al. 2011) where monaural instantaneous frequency glides in BM could account for characteristic features of binaural ITD processing. For ICIs larger than 5 ms (e.g., for an ICI of 8 ms in the current paper), where no peripheral suppression occurs, central (binaural) processes are likely responsible for raising the lateralization thresholds. Furthermore, the comparison of monaural and binaural behavioral echo thresholds (Fig. A.2a) did not show any contribution of binaural processes to fusion, in agreement with previous studies (Litovsky et al. 1997; Rakerd et al. 1997), suggesting that binaural processes might not be involved in echo threshold determination.

A.5 Conclusion

In conclusion, the results of the present study show a correlation between mechanical cochlear processes and psychoacoustical measures of the PE for short ICIs. Although low-level effects cannot be sufficient to account for all aspects of precedence, experimental evidence was provided that monaural peripheral suppression plays a fundamental role for the binaurally perceived PE for short lead-lag delays (i.e., 1-4ms). Not only do BM lead-lag interactions occur within the same time range as the behaviorally determined precedence window for clicks, they also represent the main component of lag suppression at the level of the auditory brainstem. The findings of the present study apply for click stimuli. For stimuli of longer duration than clicks, inhibitory processes may account for some aspects of the PE (Braasch and Blauert 2003; Lindemann 1986; Xia et al. 2010). Longer durations

of suppression (above 5 ms) may be explained by central processes occurring at stages above the brainstem (Blauert 1997; Damaschke et al. 2005; Liebenthal and Pratt 1999; Sanders et al. 2008).

Acknowledgments

This work was supported by the Technical University of Denmark and the Oticon Foundation.

A.6 References

- Altman, D.G., Bland, J.M. (2011) "How to obtain the p value from a confidence interval". In *BMJ* 343:d2304.
- Blauert, J. (1997) "Spatial hearing: the psychophysics of human sound localization". The MIT Press, Cambridge, MA.
- Blauert, J., Divenyi, P.L. (1988) "Spectral selectivity in binaural contralateral inhibition". In *Acustica* 66, pp. 267–274.
- Braash, J., Blauert, J. (2003) "The precedence effect for noise bursts of different bandwidths. II. Comparison of model algorithms". In *Acoust Sci Tech* 24.5: pp. 293–303.
- Damschke, J., Riedel, H., Kollmeier, B. (2005) "Neural correlates of the precedence effect in auditory evoked potentials". In *Hear Res* 205: pp. 157–171.
- Dau, T., Wagner, O., Mellert, V., Kollmeier, B. (2000) "Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion". In *J Acoust Soc Am* 107: pp. 1530–1540.
- Divenyi, P.L. (1992) "Binaural suppression of nonechoes". In *J Acoust Soc Am* 91: pp. 1078–1084.

- Divenyi, P.L., Blauert, J. (1987) "On creating a precedent for binaural patterns: when is an echo an echo?" In *Auditory processing of complex sounds*. Erlbaum, Hillsdale, NJ, pp. 146–155.
- Dizon, R.M., Colburn, H.S. (2006) "The influence of spectral, temporal, and interaural stimulus variations on the precedence effect". In *J Acoust Soc Am* 119: pp. 2947–2964.
- Fisher, B.J., Steinberg, L.J., Fontaine, B., Brette, R., Peña, J.L. (2011) "Effect of instantaneous frequency glides on interaural time difference processing by auditory coincidence detectors". In *PNAS* 108.44: pp. 18138–18143.
- Fitzpatrick, D.C., Kuwada, S., Batra, R., Trahiotis, C. (1995) "Neural responses to simple simulated echoes in the auditory brainstem of the unanesthetized rabbit". In *J Neurophys* 74.6: pp. 2469–2486.
- Fitzpatrick, D.C., Kuwada, S., Kim, D.O., Parham, K., Batra, R. (1999) "Responses of neurons to click-pairs as simulated echoes: auditory nerve to auditory cortex". In *J Acoust Soc Am* 106: pp. 3460–3472.
- Freyman, R.L., Klifton, R.K., Litovsky, R.Y. (1991) "Dynamic processes in the precedence effect". In *J Acoust Soc Am* 90: pp. 874–884.
- Goblick, T. Jr., Pfeiffer, R. (1969) "Time-domain measurements of cochlear nonlinearities using combination click stimuli". In *J Acoust Soc Am* 46.2: pp. 924–938.
- Harte, J.M., Pigasse, G., Dau, T. (2009) "Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses". In *J Acoust Soc Am* 126.3: pp. 1291–1301.
- Hartung, K., Trahiotis, C. (2001) "Peripheral auditory processing and investigations of the precedence effect which utilize successive transient stimuli". In *J Acoust Soc Am* 110: pp. 1505–1513.

- Jedrzejczak, W.W., Blinowska, K.J., Konopka, W. (2005) "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise". In *Hear Res* 205: pp. 249–255.
- Junius, D., Dau, T. (2005) "Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses". In *Hear Res* 205: pp. 53–67.
- Kapadia, S., Lutman, M. (2000) "Nonlinear temporal interactions in click-evoked otoacoustic emissions. Experimental data". In *Hear Res* 146: pp. 89–100.
- Kemp, D., Chum, R. (1980 a) "Properties of the generator of stimulated acoustic emissions". In *Hear Res* 2: pp. 213–232.
- Kemp, D., Chum, R. (1980 b) "Observations on the generator mechanism of stimulus frequency otoacoustic emissions, two-tone suppression". In: Brink G, Bilsen FA (eds) *Psychophysical, physiological and behavioural studies in hearing*. Delft University Press, Delft, pp. 34–41.
- Liebenthal, E., Pratt, H. (1999) "Human auditory cortex electrophysiological correlates of the precedence effect: binaural echo lateralization suppression". In *J Acoust Soc Am* 106: pp. 291–303.
- Lindemann, W. (1986) "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wavefront". In *J Acoust Soc Am* 80: pp. 1623–1630.
- Litovsky, R.Y., Yin, T.C. (1998) "Physiological studies of the precedence effect in the inferior colliculus of the cat". In *J Neurophysiol* 80: pp. 1302–1316.
- Litovsky, R.Y., Hawley, M.L., Colburn, H.S. (1997) "Measurement of precedence in monaural listeners". In *Meeting of the American Speech and Hearing Association*, Boston, MA.
- Litovsky, R.Y., Colburn, H.S., Yost, W., Guzman, S.J. (1999) "The precedence effect". In *J Acoust Soc Am* 106: pp. 1633–1654.

- Meddis, R. (1986) "Simulation of mechanical to neural transduction in the auditory receptor". In *J Acoust Soc Am* 79: pp. 702–711.
- Moleti, A., Sisto, R. (2008) "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions". In *J Acoust Soc Am* 123.3: pp. 1495–1503.
- Moleti, A., Sisto, R., Paglialonga, A., Sibella, E., Anteunis, L., Parazzini, M., Tognola, G. (2008) "Transient evoked otoacoustic emission latency and estimates of cochlear tuning in preterm neonates". In *J Acoust Soc Am* 124.5: pp. 2984–2994.
- Neeb, S.T., Norton, S.J., Gorga, M.P., Jesteadt, W. (1988) "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli". In *J Acoust Soc Am* 83.2: pp. 652–656.
- Parham, K., Zhao, H.B., Kim, D.O. (1996) "Responses of auditory nerve fibers of the unanesthetized decerebrate cat to click pairs as simulated echoes". In *J Neurophys* 76.1: pp. 17–29.
- Patterson, R., Allerhand, M., Giguere, C. (1995) "Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform". In *J Acoust Soc Am* 98: pp. 1890–1894.
- Picton, T.W. (2011) *Human Auditory Evoked Potentials*, Plural Publishing.
- Puria, S. (2003) "Measurements of human middle ear forward and reverse acoustics: implications for otoacoustic emissions". In *J Acoust Soc Am* 113.5: pp. 2773–2789.
- Rakerd, B., Hsu, J., Hartmann, W.M. (1997) "The Haas effect with and without binaural differences". In *J Acoust Soc Am* 101.5: p. 3083.
- Sanders, L.D., Joh, A.S., Keen, R.E., Freyman, R.L. (2008) "One sound or two? Object-related negativity indexes echo perception". In *Percept Psychophys* 70.8: pp. 1558–1570.

- Shera, C.A., Guinan, J.J. Jr (1999) "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs". In *J Acoust Soc Am* 105: pp. 782–798.
- Shera, C.A., Guinan, J.J. Jr, Oxenham, A.J. (2002) "Revised estimates of human cochlear tuning from otoacoustic and behavioral methods". In *PNAS* 99.5: pp. 3318–3323.
- Shinn-Cunningham, B.G., Zurek, P.M., Durlach, N.I., Clifton, R.K. (1995) "Cross-frequency interactions in the precedence effect". In *J Acoust Soc Am* 98: pp. 164–171.
- Tollin, D.J. (1998) "Computational model of the lateralization of clicks and their echoes". In: S. Greenberg and M. Slaney (eds) In *Proceedings of the NATO Advanced Study Institute on Computational Hearing*: pp. 77–82.
- Tollin, D.J., Henning, G.B. (1998) "Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect". In *J Acoust Soc Am* 104: pp. 3030–3038.
- Tollin, D.J., Henning, G.B. (1999) "Some aspects of the lateralization of echoed sound in man. II. The role of the stimulus spectrum". In *J Acoust Soc Am* 105: pp. 838–849.
- Tollin, D.J., Populin, L.C., Yin, T.C.T. (2004) "Neural correlates of the precedence effect in the inferior colliculus of behaving cats". In *J Neurosci* 92.6: pp. 3286–3297.
- Verhulst, S., Harte, J., Dau, T. (2011A) "Temporal suppression of the click-evoked otoacoustic emission level-curve". In *J Acoust Soc Am* 129: pp. 1452–1463.
- Verhulst, S., Shera, C.A., Harte, J., Dau, T. (2011) "Can a static nonlinearity account for the dynamics of otoacoustic emission suppression?". In: C. Shera and E. Olson (eds) In *What fire is in mine ears: progress in*

auditory biomechanics. Proceedings of the 11th International Mechanics of Hearing Workshop, American Institute of Physics. pp. 257–263

- Verhulst, S., Bianchi, F., Dau, T. (2013) "Cochlear contributions to the precedence effect". In: Moore B, Patterson R, Winter I, Carlyon R, Gockel H (eds) In *Basic aspects of hearing: physiology and perception*. Springer, New York.
- Wallach, H., Newman, E.B., Rosenzweig, M.R. (1949) "The precedence effect in sound localization". In *Am J Psychol* 62: pp. 315–336.
- Wickesberg, R.E., Oertel, D. (1990) "Delayed, frequency-specific inhibition in the cochlear nuclei of mice: a mechanism for monaural echo suppression". In *J Neurosci* 10.6: pp. 1762–1768.
- Wolf, M., Schuchmann, M., Wiegrebe, L. (2010) "Localization dominance and the effect of frequency in the Mongolian gerbil", *Meriones unguiculatus*. In *J Comp Physiol* 196: pp. 463–470.
- Xia, J., Shinn-Cunningham, B.G. (2011) "Isolating mechanisms that influence measures of the precedence effect: theoretical predictions and behavioral tests". In *J Acoust Soc Am* 130: pp. 866–882.
- Xia, J., Brughera, A., Colburn, H.S., Shinn-Cunningham, B.G. (2010) "Physiological and psychophysical modeling of the precedence effect". In *J Assoc Res Otolaryngol* 11: pp. 495–513.
- Yang, X., Grantham, D.W. (1997A) "Echo suppression and discrimination suppression aspects of the precedence effect". In *Percept Psychophys* 59: pp. 1108–1117.
- Yang, X., Grantham, D.W. (1997B) "Cross-spectral and temporal factors in the precedence effect: discrimination suppression of the lag sound in free-field". In *J Acoust Soc Am* 102: pp. 2973–2983.
- Zurek, P.M. (1980) "The precedence effect and its possible role in the avoidance of interaural ambiguities". In *J Acoust Soc Am* 67: pp. 952–964.

-
- Zurek, P.M. (1987) "The precedence effect". In: W. A. Yost and G. Gourevitch (eds) In *Directional hearing*. Springer, New York. pp. 85–105.
- Zweig, G., Shera, C.A. (1995) "The origin of periodicity in the spectrum of evoked otoacoustic emissions". In *J Acoust Soc Am* 98: pp. 2018–2047.

Contributions to Hearing Research

- Vol. 1:** *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.
- Vol. 2:** *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.
- Vol. 3:** *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.
- Vol. 4:** *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.
- Vol. 5:** *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.
- Vol. 6:** *Helen Connor*, Hearing aid amplification at soft input levels, 2010.
- Vol. 7:** *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.
- Vol. 8:** *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.
- Vol. 9:** *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.
- Vol. 10:** *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.

- Vol. 11:** *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.
- Vol. 12:** *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.
- Vol. 13:** *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.
- Vol. 14:** *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.
- Vol. 15:** *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.
- Vol. 16:** *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.
- Vol. 17:** *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.
- Vol. 18:** *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.
- Vol. 19:** *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.
- Vol. 20:** *Michał Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
- Vol. 21:** *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.
- Vol. 22:** *Federica Bianchi*, Complex-tone pitch representations in the human auditory system, 2016.

The end.

To be continued...

Understanding how the human auditory system processes the physical properties of an acoustical stimulus to give rise to a pitch percept is a fascinating aspect of hearing research. Since most natural pitch-evoking sounds are harmonic complex tones, this work focused on the nature of pitch-relevant cues that are necessary for the auditory system to retrieve the pitch of complex sounds. The existence of different pitch-coding mechanisms for resolved and unresolved harmonics was investigated by comparing pitch-discrimination performance across different cohorts of listeners, specifically those showing enhanced pitch cues (i.e., musicians) and those typically having disrupted pitch cues (i.e., hearing-impaired listeners). In particular, two main topics were addressed: the relative importance of resolved and unresolved harmonics for normal-hearing and hearing-impaired listeners and the effect of musical training for pitch discrimination of resolved and unresolved harmonics. Overall, the findings presented throughout this work strengthen the hypothesis that loss of cochlear compression leads to an enhanced envelope representation of the unresolved harmonics in hearing-impaired listeners relative to normal-hearing listeners. Additionally, behavioral and objective investigations in musicians and non-musicians suggest the existence of a training-dependent effect in the musicians that is partly specific to the resolvability of the harmonics.

DTU Electrical Engineering

Department of Electrical Engineering

Ørsteds Plads
Building 348
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 38 00
Fax: (+45) 45 93 16 34
www.elektro.dtu.dk