Neural coding and perception of pitch in the normal and impaired human auditory system

Santurette, Sébastien; Dau, Torsten; Buchholz, Jörg; Wouters, Jan; Oxenham, Andrew J.

Publication date:
2011

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

Link back to DTU Orbit

Citation (APA):
Santurette, S., Dau, T., Buchholz, J., Wouters, J., & Oxenham, A. J. (2011). Neural coding and perception of pitch in the normal and impaired human auditory system. (Contributions to hearing research; No. 10).

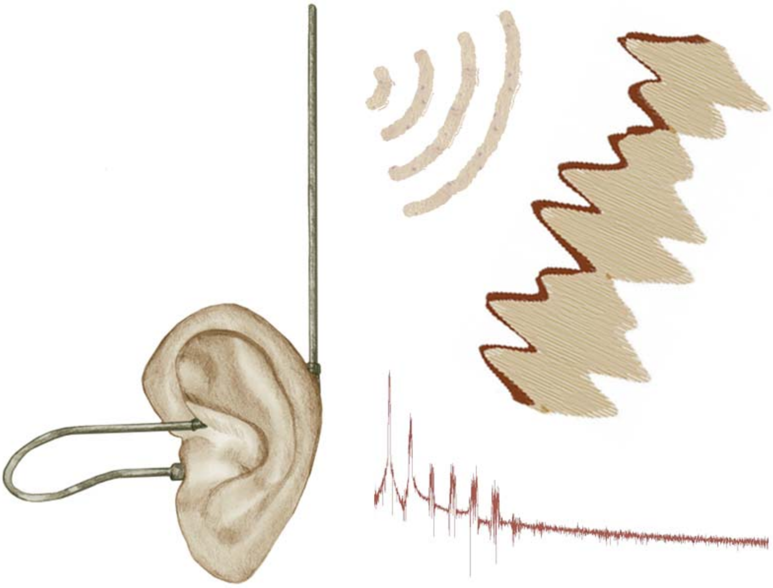
Technical University of Denmark

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 10

Sébastien Santurette

Neural coding and perception of pitch in the normal and impaired human auditory system



Centre for Applied Hearing Research
Department of Electrical Engineering

Neural coding and perception of pitch in the normal and impaired human auditory system

PhD thesis by
Sébastien Santurette

CAHR
Centre for Applied Hearing Research

Technical University of Denmark
2011

Cover illustration: *Earamin* by Emily Allen.
Printed in Denmark by Rosendahls - Schultz Grafisk A/S.
The defense was held on June 17, 2011.

This PhD-dissertation is the result of a research project at the Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark (Kgs. Lyngby, Denmark). Parts of the project were carried out at the Division of Experimental Otorhinolaryngology, Department of Neurosciences, Katholieke Universiteit Leuven (Leuven, Belgium), and at the Auditory Perception and Cognition Lab, Department of Psychology, University of Minnesota (Minneapolis, USA).

The project was financed by a DTU stipend. The external subprojects were supported by grants from the Fund for Scientific Research Flanders (K.U. Leuven, G.0331.08), the Research Council of Katholieke Universiteit Leuven (K.U. Leuven, OT/07/034), and the U.S. National Institute of Health (U of M, R01 DC 05216). The external stay at the University of Minnesota was further supported by travel grants from the Otto Mønsted Fund and the Oticon Foundation.

Supervisors

Main supervisor

Prof. Torsten Dau
Centre for Applied Hearing Research
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Co-supervisor

Ass. Prof. Jörg M. Buchholz
Centre for Applied Hearing Research
Department of Electrical Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

External advisors

Prof. Jan Wouters
Division of Experimental ORL
Department of Neurosciences
Katholieke Universiteit Leuven
Leuven, Belgium

Prof. Andrew J. Oxenham
Auditory Perception and Cognition Lab
Department of Psychology
University of Minnesota
Minneapolis, MN, USA

Abstract

Pitch is an important attribute of hearing that allows us to perceive the musical quality of sounds. Besides music perception, pitch contributes to speech communication, auditory grouping, and perceptual segregation of sound sources. In this work, several aspects of pitch perception in humans were investigated using psychophysical methods. First, hearing loss was found to affect the perception of binaural pitch, a pitch sensation created by the binaural interaction of noise stimuli. Specifically, listeners without binaural pitch sensation showed signs of retrocochlear disorders. Despite adverse effects of reduced frequency selectivity on binaural pitch perception, the ability to accurately process the temporal fine structure (TFS) of sounds at the output of the cochlear filters was found to be essential for perceiving binaural pitch. Monaural TFS processing also played a major and independent role for a variety of basic auditory tasks, indicating that it may be a crucial measure to consider for hearing-loss characterization. In contrast to hearing-impaired listeners, adults with dyslexia showed no deficits in binaural pitch perception, suggesting intact low-level auditory mechanisms. The second part of this work investigated the role of temporal and spectral information for complex pitch perception. In particular, it was shown that the low pitch evoked by high-frequency complex tones was not conveyed by temporal-envelope cues as such. Moreover, the fact that the individual frequency components could not be heard out separately by the listeners suggested that the low pitch relied on TFS information, even in high-frequency regions where phase-locking in auditory-nerve cells is believed to be weak. A second set of experiments could however not validate the assumption of a temporally-coded pitch and indicated that the use of spectral cues remained plausible. Simulations of auditory-nerve representations of the complex tones further suggested that a spectrotemporal mechanism combining precise timing information across auditory channels might best account for the behavioral data. Overall, this work provides insights into the fundamental auditory mechanisms underlying pitch perception, and may have implications for future pitch-perception models, as well as strategies for auditory-profile characterization and restoration of accurate pitch perception in impaired hearing.

Resumé

Tonehøjde er en vigtig egenskab ved hørelsen, der giver os mulighed for at opfatte musikalsk kvalitet af lyde. Udover opfattelse af musik, bidrager tonehøjde til talekommunikation, auditiv gruppering, og perceptuel adskillelse af lydkilder. I denne afhandling blev adskillige aspekter af tonehøjde i det menneskelige auditive system undersøgt ved hjælp af psykofysiske forsøg. Først sås det at høretab påvirker opfattelsen af binaural pitch, en opfattelse af tonehøjde skabt af den binaurale interaktion af støjstimuli. Især viste individer uden binaural-pitch opfattelse tegn på retrocochlear lidelser. Evnen til præcist at bearbejde lydes temporale finstruktur (TFS) ved udgangen af det indre øres filtre sås at være af afgørende betydning også selv om de negative effekter ved reduceret frekvensselektivitet tages i betragtning. Bearbejdning af den monaurale TFS sås også at spille en vigtig og selvstændig rolle for en række grundlæggende auditive funktioner, hvilket tyder på, at denne evne kunne være en vigtig størrelse at måle for karakterisering af høretab. I modsætning til hørehæmmede viste personer med ordblindhed intet besvær med at opfatte binaural pitch, hvilket antyder intakte lavniveau auditive mekanismer. Den anden del af dette arbejde undersøgte rollen af temporal og spektral information for komplekse toners tonehøjde. Mest bemærkelsesværdigt blev det påvist, at tonehøjden for tonekomplekser med høje frekvenser ikke blev bestemt ud fra deres temporale indhylningskurve. Desuden kunne de enkelte frekvenskomponenter ikke adskilles, hvilket antyder at tonehøjden bestemmes fra TFS oplysninger, selv i højfrekvens regioner, hvor synkronien i auditive nerveceller menes at være svag. Andre forsøg kunne dog ikke bekræfte antagelsen om en temporalt kodet tonehøjde og viste, at muligheden for spektral kodning foreligger. Simuleringer af stimuliene i den auditive nerve tydede endvidere på, at en spektrotemporal mekanisme, der kombinerer den temporale information på tværs af auditive kanaler, muligvis bedst stemmer overens med de adfærdsmæssige data. Samlet set giver dette arbejde indsigt i de grundlæggende auditive mekanismer bag opfattelsen af tonehøjde, og kan have betydning for fremtidige auditive modeller, samt strategier for karakterisering af den auditive profil og retablering af tonehøjde opfattelse hos personer med nedsat hørelse.

Résumé

La hauteur tonale est un élément important de l'audition nous permettant de percevoir la qualité musicale des sons. Outre la perception musicale, la hauteur facilite la communication orale, le groupement auditif, et la ségrégation de sources sonores. Par le biais d'expériences psychophysiques, cette thèse étudie plusieurs aspects de la perception de la hauteur dans le système auditif humain. Tout d'abord, il est démontré que la perte d'audition affecte la perception du pitch binaural, une sensation de hauteur tonale créée par l'interaction binaurale de deux bruits. Plus précisément, des signes de troubles rétrocochléaires sont trouvés chez les sujets sans sensation de pitch binaural. Bien que la perception du pitch binaural soit affectée par une sélectivité en fréquence réduite, la capacité de traiter la structure temporelle fine (STF) des sons à la sortie des filtres cochléaires s'impose en facteur essentiel. De plus, les résultats indiquent un rôle majeur et indépendant du traitement périphérique de la STF pour une variété de fonctions auditives de base, suggérant qu'une telle mesure pourrait être importante à inclure dans des tests de caractérisation de la perte auditive. Contrairement aux malentendants, il est observé que les adultes souffrant de dyslexie ne présentent aucun déficit dans la perception du pitch binaural, suggérant des mécanismes auditifs de bas niveau intacts. La deuxième partie de ce travail s'intéresse au rôle de l'information temporelle et spectrale pour la perception de la hauteur de tons complexes. En particulier, il est montré que la hauteur de tons complexes à fréquences élevées n'est pas déterminée par leur enveloppe temporelle en tant que telle. De plus, la difficulté des sujets à extraire les composantes spectrales individuelles suggère que la hauteur perçue est basée sur l'information contenue dans la STF, même dans les régions de haute fréquence où la synchronie des neurones auditifs est supposée faible. Une deuxième série d'expériences n'est cependant pas en mesure de valider l'hypothèse d'un code temporel et indique que l'utilisation d'indices spectraux reste plausible. En outre, une simple simulation de représentations internes des stimuli au niveau du nerf auditif suggère qu'un mécanisme combinant l'information temporelle entre les différents canaux auditifs pourrait mieux refléter les données expérimentales. Dans l'ensemble, ce travail permet de mieux comprendre les mécanismes sous-jacents à la perception de la hauteur tonale, avec de potentielles implications pour de futurs modèles de l'audition, ainsi que pour des stratégies de caractérisation du profil auditif et de restauration de la sensation de hauteur tonale chez les malentendants.

Preface

"I do not ever want to become a researcher." I remember pronouncing this sentence ten years ago, after a failed attempt to enter a renowned French research institution. My, these "researchers" scared me! They dressed funny, worked in basements, and talked about things I did not understand. Or could it be because they were better known to me as "examiners" at that time? Whatever the reason, I surely would not want to become one of them! A few years in Denmark, pushing the doors of building 352 in a town bearing the awkward name of Kgs. Lyngby, and something strange happened... Well, I guess only fools do not change their minds!

These years as a PhD student have been an enriching experience. The people who have surrounded me during this adventure have nevertheless, from my point of view, been at the center of it all. I would thus like to thank:

Torsten Dau for his guidance and great support throughout the project. His "Auditory signal processing and perception" course was decisive in my choice of specialization in psychoacoustics. Torsten, I sincerely thank you for trusting me and giving me a lot of freedom in the evolution of this work. It has been a pleasure to work with you, and your dedication and enthusiasm for this field of research have given me much motivation and inspiration.

The staff and students at CAHR and Acoustic Technology, for creating such a friendly and stimulating work and study environment. Your support as colleagues and friends has been essential throughout these years. Thanks also go to all of you who offered their help concerning the scientific, technical, or administrative aspects of the project.

All researchers who I had the opportunity to share my work with. I am grateful to all of you who took some time for scientific discussions during your visit at CAHR or at meetings and conferences. I have learned a lot from your comments and suggestions, which have played an essential role in many aspects of this work. Many thanks also go to the different people who gave me the chance to present my work at seminars and workshops.

All test persons who were involved in this research and volunteered to spend some time in the listening booth, some of them for many hours. Without your invaluable contribution, this work would not have been possible.

Jan Wouters, for welcoming me in Leuven for several short stays. I am very grateful to you, Hanne Poelmans, Heleen Luts, and other colleagues from the ExpORL lab, for your help throughout this friendly and fruitful collaboration.

Andrew J. Oxenham, for his kind hospitality. Thank you for making my three months in Minneapolis comfortable research-wise, music-wise, and social-wise, and for introducing me to the art of understanding the rules of baseball.

All researchers and students at the APC lab for enjoyable discussions. I had a great time sharing thoughts with all of you about the highly important topics that are movies, English vocabulary, coffee-making, reality, truth, and psychoacoustics.

Emily Allen, for her contribution to the cover illustration of this book.

Thomas Ulrich Christiansen, for his help with the Danish translation of the abstract.

All the great filmmakers and musicians whose art helped me keep my head above water and my feet on (and especially off!) the ground.

My family, friends, and partner, whose love and support I could not have done without.

Sébastien Santurette, March 4, 2011.

Related publications

Journal articles

- Santurette S., Dau T. (2011). Relating binaural pitch perception to the individual listener's auditory profile. *J. Acoust. Soc. Am.*, submitted.
- Santurette S., Dau T. (2011). The role of temporal fine structure information for the low pitch of high-frequency complex tones. *J. Acoust. Soc. Am.*, 129(1):282–292.
- Santurette S., Poelmans H., Luts H., Ghesquière P., Wouters J., Dau T. (2010). Detection and identification of monaural and binaural pitch contours in dyslexic listeners. *J. Assoc. Res. Otoloaryngol.*, 125(11):515–524.
- Santurette S., Dau T. (2007). Binaural pitch perception in normal-hearing and hearing-impaired listeners. *Hear. Res.*, 223(1-2):29–47.

Conference papers

- Santurette S., Dau T., Poelmans H., Luts H., Wouters J. (2010). Detection and identification of monaural and binaural pitch contours in dyslexic listeners. *International Symposium on Auditory and Audiological Research (ISAAR), Binaural Processing and Spatial Hearing*, Helsingør, Denmark, August 2009.
- Dau T., Santurette S., Strelcyk, O. (2007). Binaural pitch perception in hearing-impaired listeners. *Forschritte der Akustik DAGA'07, 33rd German Convention on Acoustics*, Stuttgart, Germany, March 2007.

Published abstracts

- Santurette S., Dau T. (2010). Behavioral Measures of Monaural Temporal Fine Structure Processing. *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe, CA, August 2010.
- Santurette S., Dau T. (2010). The Low Pitch of High-Frequency Complex Tones Relies on Temporal Fine Structure Information. *Association for Research in Otolaryngology (ARO), 33rd Mid-Winter Meeting*, Anaheim, CA, February 2010.
- Santurette S., Dau T. (2009). Processing of Binaural Pitch Stimuli in Hearing-Impaired Listeners. *Association for Research in Otolaryngology (ARO), 32nd Mid-Winter Meeting*, Baltimore, MD, February 2009.
- Santurette S., Dau T. (2008). Relating the absence of binaural pitch percept to retro-cochlear impairment. *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe, CA, August 2008.
- Santurette S., Dau T. (2006). Melody recognition with binaural-pitch stimuli in normal-hearing and hearing-impaired listeners. *Association for Research in Otolaryngology (ARO), 29th Mid-Winter Meeting*, Baltimore, MD, February 2006.

Contents

List of abbreviations

*n*I-*m*AFC	*n*-interval, *m*-alternative, forced-choice
ACF	Autocorrelation function
ALT	Alternating sine-cosine phase
AM	Amplitude modulation
ANOVA	Analysis of variance
BBN	Broadband noise
BBP	Binaural band pitch
BEP	Binaural edge pitch
BIBP	Binaural incoherent-band pitch
BICEP	Binaural coherence edge pitch
BILD	Binaural intelligibility level difference
BM	Basilar membrane
BMCL	Binaural most comfortable level
BMLD	Binaural masking level difference
BP	Binaural pitch
CAP	Central activity pattern
CEOAE	Click-evoked otoacoustic emission
CF	Characteristic frequency
CT	Combination tone
cu	Categorical unit
EC	Equalization-cancellation
EM	Expectation-maximization

ERB	Equivalent rectangular bandwidth
ERB-As.	Equivalent rectangular bandwidth asymmetry
EWH	Equal-weight histogram
F0	Fundamental frequency
FFT	Fast Fourier transform
FIR	Finite-impulse-response
FM	Frequency modulation
FM-As.	Frequency-modulation detection threshold asymmetry
FMDT	Frequency-modulation detection threshold
HF	High-frequency
HI	Hearing-impaired
H/I	Harmonic/inharmonic
HL	Hearing level
HP	Huggins’ pitch
HRTF	Head-related transfer function
iFFT	Inverse fast Fourier transform
ILD	Intelligibility level difference
IPD	Interaural phase difference
ISI	Inter-spike interval
JND	Just noticeable difference
LEC	Lowest effective component
LF-HL	Low-frequency hearing loss
LNN	Low-noise noise
MASD	Mean absolute spatial derivative
MCL	Most comfortable level
mEC	Modified equalization-cancellation
MP	Monaurally-detectable pitch
NBN	Narrow-band noise
NH	Normal-hearing

OAE	Otoacoustic emission
OD	Obscure dysfunction
pdf	Probability density function
peSPL	Peak-equivalent sound pressure level
PM	Pitch matching
RAN	Rapid automatized naming
RBICEP	Reversed binaural coherence edge pitch
RMS	Root-mean-square
roex	Rounded-exponential
rs	Remaining subjects
SACF	Summary autocorrelation function
SD	Standard deviation
SIN	Sine phase
SISI	Short increment sensitivity index
SL	Sensation level
SNR	Signal-to-noise ratio
SPL	Sound pressure level
SRT	Speech reception threshold
TFS	Temporal fine structure
TN	Tone in noise
TS	Transposed stimulus

1

Theories and models of pitch perception: An overview

In this first chapter, the evolution of pitch perception theories is presented, as well as an overview of the main types of models that have been proposed to account for pitch perception in the normal human auditory system. The outline of the present work is then briefly described.

1.1 Introduction

Pitch is one of the most studied topics in hearing research and is densely represented in the scientific literature concerned with theories of hearing and auditory perception. This section gives some essential definitions and shortly presents what makes pitch a central topic in the field of hearing.

1.1.1 Why study pitch?

Pitch is used as a cue for our perception of a variety of ordinary sounds. In particular, pitch contributes to three important elements of hearing:

- *Music perception:* The pitch percept is inherent to the concepts of melodies and chords, and is what allows us to perceive a sound as musical;
- *Speech recognition:* Pitch can carry prosody information in language through intonation and stress. It can also carry semantic information: In tone languages,

two syllables with different pitch variations might correspond to two different words;

- *Sound source segregation:* Pitch is an essential cue for auditory grouping: If for instance a male and a female voice are heard simultaneously, we find it easy to segregate them because they have two different pitches.

It is because pitch is used extensively by the auditory system in our perception of everyday sounds that it has drawn the interest of many. Understanding pitch is a fundamental requirement in the constant quest for an answer to the question: Why do we hear what we hear?

1.1.2 Why is pitch complicated?

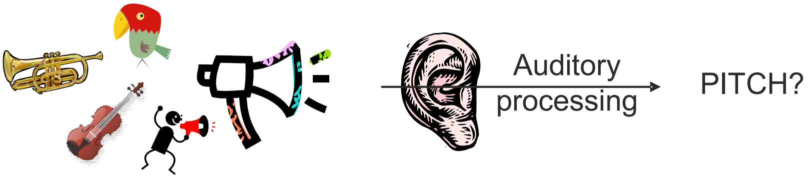
Despite an interest for pitch that dates back to ancient times and the very large amount of scientific research on the topic, mechanisms underlying pitch perception remain poorly understood. Controversy and heated debates have always been part of any progress made in understanding pitch perception, which is an indication that pitch might be a much more complex attribute than it seems at first glance. Defining pitch was already a source of disagreement, but here are what today seem to be the most widely-accepted definitions:

> “That attribute of auditory sensation in terms of which sounds might be ordered ...
> ... on a musical scale” (ASA, 1960)
> ... on a scale extending from low to high” (ANSI, 1973)

The latter definition is less restrictive than the former in the sense that it does not imply that pitch is necessarily related to music. Both definitions assume a one-dimensional perceptual scale, and that this scale is “associated with frequency” (AFNOR, 1977). More generally, the pitch or “pitch-frequency” of a sound is often defined as the frequency of a pure tone having the same pitch as that sound. It results that a sound is generally said to have a pitch, or to be a pitch-evoking stimulus, if a sinusoidal tone can be matched to it. The terms “pitch” and “pitch-frequency” are both used in the following chapters to refer to the subjective perceptual attribute, which should be distinguished from the physical properties of the pitch-evoking sound.

1.1.3 Pitch-evoking stimuli

If the concept of pitch is rather straight-forward, the question of how the auditory system processes the physical stimulus to extract its pitch is a more complicated matter.



A good starting point is to find a description of the physical stimulus. Most pitch-evoking stimuli in our environment (*e.g.*, sounds from musical instruments, animal sounds, speech) are complex harmonic sounds. Such sounds can be described as a sum of harmonic sinusoidal components:

$$\begin{cases} x(t) = \sum_{k=1}^{\infty} A_k \cos(2\pi f_k t + \varphi_k) \\ f_k = k f_1, \forall k \in \mathbb{N} \end{cases} \tag{1.1}$$

where $x(t)$ is the temporal waveform at time t, and A_k, f_k, φ_k are respectively the amplitude, frequency, and starting phase of the k^{th} component.

Such a stimulus is called a *harmonic complex*, which means that all its frequency components f_k, also called *harmonics* or *partials*, are positive integer multiples of the first component f_1. For example, the frequency of the fourth harmonic is $f_4 = 4f_1$. The frequency of the first harmonic f_1 is also called the *fundamental frequency* of the stimulus, and is usually referred to as F0. Most pitch-evoking stimuli are periodic, with a period T related to their fundamental frequency:

$$\begin{cases} x(t+T) = x(t), \forall t \geq 0 \\ T = \dfrac{1}{\mathrm{F0}} \end{cases} \tag{1.2}$$

In most cases, these stimuli elicit a pitch related to their period (thus called *periodicity pitch*), *i.e.*, their "pitch-frequency" is equal to F0. This is where the basis of pitch perception modelling lies: Most pitch perception models aim at estimating, for a given stimulus, the fundamental frequency F0, or its inverse, the period T. There are thereby two approaches to pitch perception modelling:

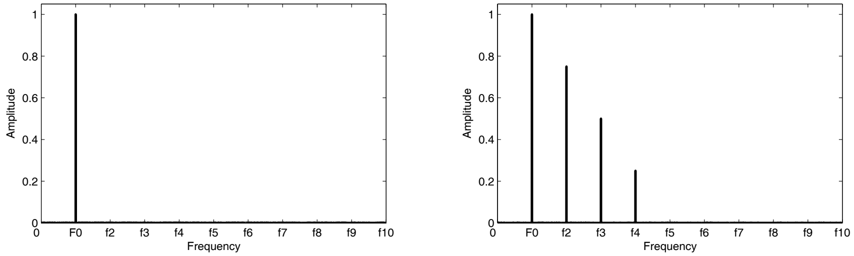
- a *spectral* approach, based on F0 estimation;
- a *temporal* approach, based on T estimation.

1.2 Spectral and temporal approaches

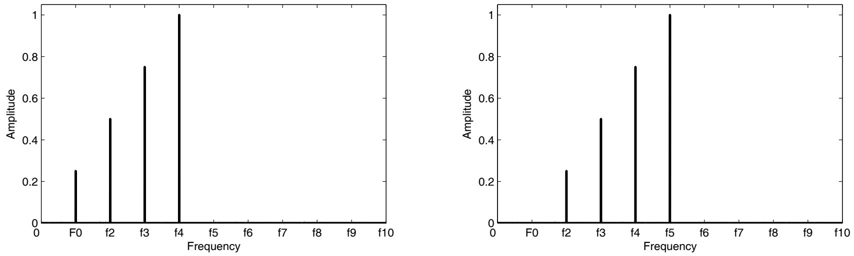
Using simple examples inspired by de Cheveigné (2005), this section introduces the two main approaches, spectral and temporal, to pitch perception modelling.

1.2.1 Spectral approach

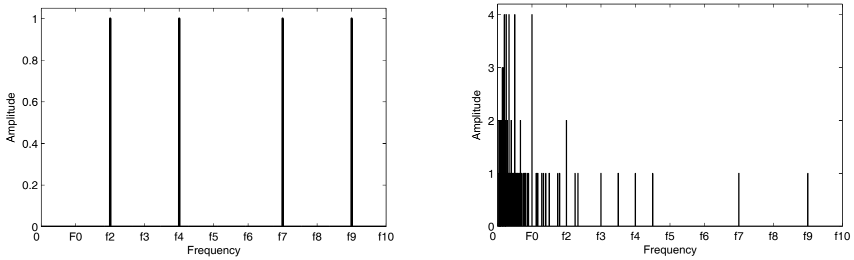
In order to estimate the fundamental frequency of a stimulus, one might start by looking at its frequency spectrum. For a sinusoidal stimulus, the spectrum might look like the one in Fig. 1.1(a). In this simple case, the peak frequency in the spectrum gives us the cue to pitch. If we now consider a stimulus with several harmonic components of different amplitudes, as in Fig. 1.1(b), it may seem reasonable to use the position of the largest spectral peak as the cue to pitch. However, if the amplitudes of the different components are modified (Fig. 1.1(c)), the pitch remains the same. We might then be tempted to think that it is not the largest peak, but the peak of lowest frequency in the spectrum, that determines the pitch. But let us now consider a stimulus with the spectrum given in Fig. 1.1(d). Despite the absence of any component at F0, the pitch still corresponds to F0. This is called the "missing fundamental" phenomenon,



(a) Pitch = 400 Hz (b) Pitch = 400 Hz



(c) Pitch = 400 Hz (d) Pitch = 400 Hz



(e) Pitch = 400 Hz or 800 Hz (f) Subharmonic histogram of (e)

Figure 1.1: Spectral approach. (a)-(e) Frequency spectra of five stimuli with F0 = 400 Hz. (f) Histogram of subharmonic frequencies derived from the spectrum in (e).

according to which a pitch at F0 might still be present, although the stimulus only contains higher harmonics. From the example in Fig. 1.1(d), one might think that the right cue for pitch is the spacing between the different partials. But this can be ruled out by the example given in Fig. 1.1(e), which has an ambiguous pitch corresponding either to F0 or f_2.

Can we then find a scheme that is able to retrieve the pitch of all these different stimuli? One solution that works for all stimuli in Fig. 1.1 is to build a histogram from the frequency spectrum. This can be done by dividing each partial by successive small integers (*i.e.*, calculate so-called *subharmonics*), and adding up one unit at the corresponding frequency each time a subharmonic shows up. With the spectrum from Fig. 1.1(e), one obtains the histogram shown in Fig. 1.1(f). The cue to pitch is then given by the rightmost largest peak in the histogram, in our case corresponding to F0. Note that the ambiguous pitch at f_2 might also be explained by a somewhat lower peak that stands out at that frequency in the histogram. Such a method is called a *pattern-matching scheme*: We have derived a pattern indexed by frequency (in this case a histogram), from which the pitch is then extracted.

1.2.2 Temporal approach

In the temporal approach, one wishes to estimate the period T of the stimulus. For that purpose, one might start by looking at the stimulus waveform. For a sinusoidal stimulus, the waveform looks like the one shown in Fig. 1.2(a). For this simple stimulus, the interval between two consecutive peaks in the waveform gives us the stimulus period T. The inverse of that period, $1/T$, corresponds to the pitch frequency. Let us now consider the waveform in Fig. 1.2(b). The pitch is the same as that of the previous sinusoidal stimulus, but the interval between two consecutive peaks does not lead to a correct period estimation, and one has to use only the largest peaks to find the right cue to pitch. With the case of Fig. 1.2(c), which shows the inverse waveform of the previous example and thus has the same pitch, one can see that it might be better to use successive positive zero-crossings for period estimation, rather than peaks in the waveform. But let us now consider another example eliciting the same pitch, shown in Fig. 1.2(d). Here, zero-crossings cannot be used for period estimation. However, the stimulus has a clearly periodic temporal envelope (dotted curve). The interval between two peaks in the temporal envelope is equal to T and can thus serve as a cue to pitch. However, it is possible to find a stimulus with the same temporal envelope that elicits a different pitch, as the one shown in Fig. 1.2(e). Here, the pitch frequency is not equal to the inverse of the envelope period $1/T_{env}$. It thus seems that envelope period is not a satisfactory cue to pitch.

Can we once again find a scheme that would retrieve the pitch of all these stimuli? A way to do so is to compare each sample of the stimulus to every other sample in turn and detect inter-sample intervals with a good match. This can be achieved by a simple mathematical tool, the *autocorrelation* function r. Autocorrelation consists in multiplying the signal $x(t)$ by a delayed version of itself, for a given time-delay τ, then integrating over time:

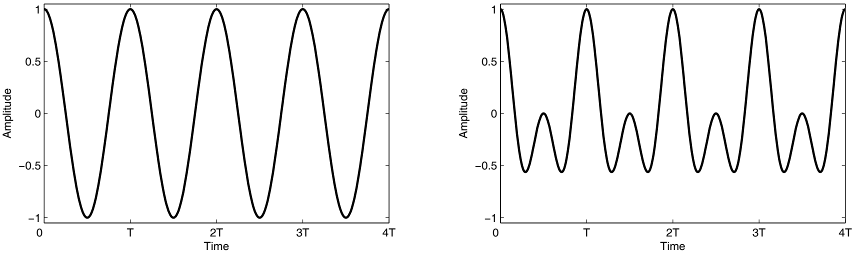
$$r(\tau) = \int x(t)x(t-\tau)dt \tag{1.3}$$

Fig. 1.2(f) shows the autocorrelation function of the stimulus from Fig. 1.2(e). The delay at the first highest-amplitude maximum in the function gives a correct estimation of the stimulus period T. In this specific case, the presence of maxima of large amplitude around $T_{env} = T/2$ can explain the perception of an ambiguous pitch related to the inverse of the correponding delays.

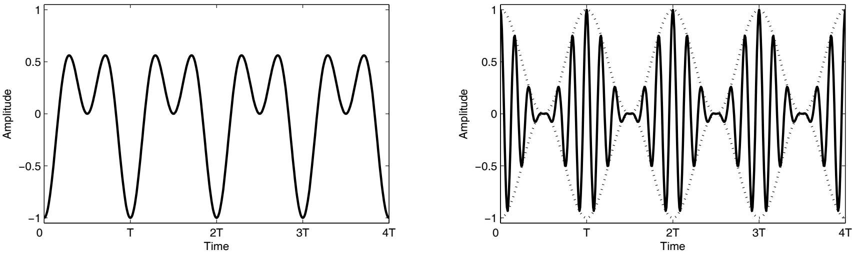
1.2.3 Two theories of pitch

We have just presented two approaches for the extraction of the pitch of a given stimulus:

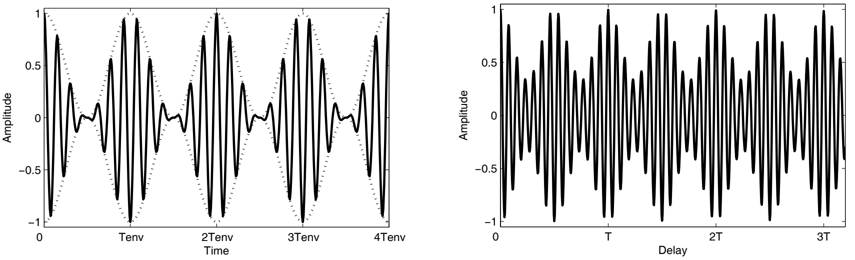
- A *spectral* approach, in which *pattern-matching* schemes seem to be efficient for correct pitch determination. Pattern-matching models are based on the *place theory* of pitch perception. According to the place theory, selective frequency analysis and tonotopicity in the auditory system are the key elements for pitch extraction;
- A *temporal* approach, in which *autocorrelation* schemes seem to be efficient for correct pitch determination. Autocorrelation models are based on the *time*



(a) Pitch = 400 Hz (b) Pitch = 400 Hz



(c) Pitch = 400 Hz (d) Pitch = 400 Hz



(e) Pitch $\neq$ 400 Hz (f) Autocorrelation function of (e)

Figure 1.2: Temporal approach. (a)-(d) Temporal waveforms of five stimuli with $T = 1/400 = 0.0025$ s. (e) Waveform of a stimulus with an envelope period $T_{env} = 0.0025$ s. (f) Autocorrelation function of the stimulus in (e).

theory of pitch perception. According to the time theory, precise temporal coding in the auditory nerve is the key element for pitch extraction.

The following section describes in more details these two fundamental theories of pitch perception.

1.3 Theories and models of pitch perception

Whether the extraction of pitch by the auditory system relies on place or time information is at the center of a historical and still ongoing controversy (see, *e.g.*, de Boer (1976) and de Cheveigné (2005) for reviews). The origins and evolutions of the place and time theories of pitch perception are presented in this section, as well as the main types of pitch perception models they gave rise to.

1.3.1 Place theory

Introduction

According to the place theory of pitch perception, frequency information transmitted along the auditory pathway is exclusively used by the system to derive the pitch of a given stimulus. As illustrated in Fig. 1.3, the input sound has different frequency components, which are assumed to stimulate corresponding places along the basilar membrane. In turn, auditory-nerve fibers with corresponding characteristic frequencies are excited and produce neural spikes. It is then these characteristic frequencies and the amount of activation in each channel that are used by the system to extract pitch.

Historical background

The origins of the place theory date back to antique times, when Greek mathematicians found a relationship between string length ratios and musical intervals (Pythagoras, 6th century B.C.) and gave a first definition of interval and pitch (Aristoxenos, 4th century B.C.). Their finding of a qualitative dependency of pitch on frequency of vibration were supplemented by quantitative relationships in the 17th century, when Mersenne

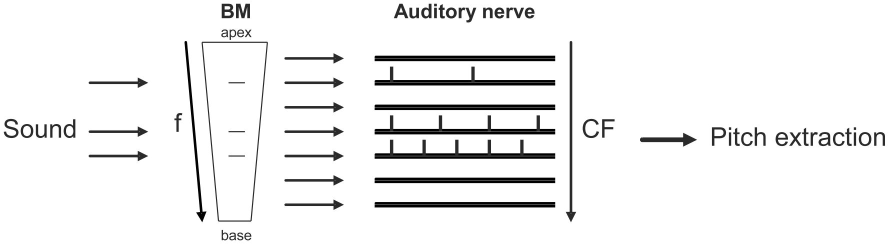


Figure 1.3: Simplified illustration of the place theory of pitch perception. f: frequency. BM: Basilar Membrane. CF: Characteristic Frequency.

(1636) and Galilei (1638) established the laws of strings and derived the frequencies of musical notes.

In the 19th century, the place theory became the dominant theory of pitch. In opposition to Seebeck (1841), who claimed that pitch was always equal to the stimulus period, regardless of the stimulus shape, Ohm (1843) and Helmholtz (1877) relied on frequency analysis, based on the findings of Fourier (1822). They were convinced that no pitch could exist without a corresponding partial. Therefore, according to the theory of Ohm (1843) and Helmholtz (1877), only periodic vibrations containing a non-zero fundamental partial could evoke a period-related pitch. But Seebeck (1841) was first to discover that a period-related pitch could still be evoked without a fundamental partial. The dominant theory of Helmholtz (1877), that was otherwise very useful to explain the pitch of pure tones and the extraction of the frequencies of partials, was heavily challenged by this problem of the “missing fundamental”. Although Helmholtz later proposed that distortion within the ear might be at the origin of the missing fundamental phenomenon and the retrieval of the fundamental pitch, this argument ended up being put down as well: Schouten (1938) showed that a complete cancellation of the fundamental partial within the ear still left the pitch unchanged. This was later confirmed in masking experiments by Licklider (1956).

Pattern-matching models of pitch perception were found to address the problem of the missing fundamental. This is at the origin of the success of this type of models.

Pattern matching

Pattern-matching models consist of two successive entities (Fig. 1.4):

- A *frequency analyzer*, that extracts the individual frequency components of the input stimulus. For each different model, a specific algorithm then derives a pattern, indexed by frequency, using these resolved frequency components.
- A *pattern recognizer*, that matches the obtained pattern to preexisting templates of harmonic frequencies assumed to be present in the system's memory.

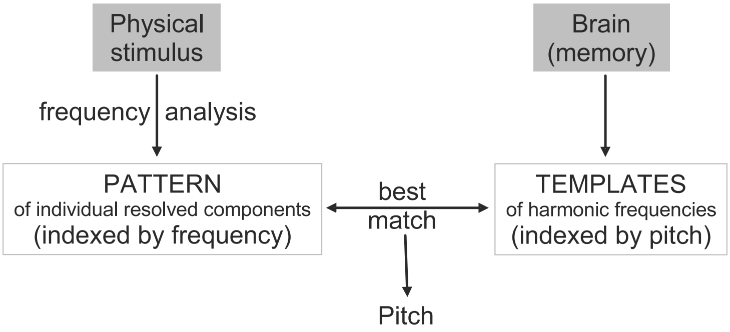
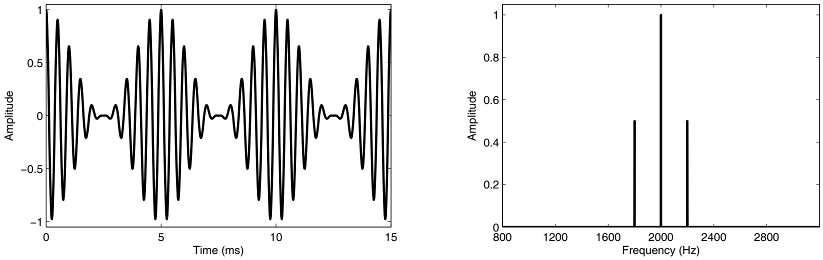


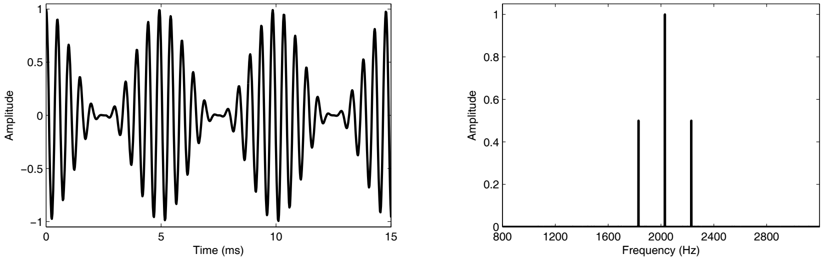
Figure 1.4: Simplified functional scheme of a pattern-matching model.

Let us consider the example of an amplitude modulated sine wave with a carrier frequency $f_c = 2000$ Hz and a modulation frequency $f_m = 200$ Hz. The waveform of such a stimulus is plotted in Fig. 1.5(a). It produces a pitch corresponding to 200 Hz. The first step of a pattern-matching model would be a frequency analysis that retrieves the three frequency components of the stimulus: 1800, 2000, and 2200 Hz, as it can be seen in the spectrum in Fig. 1.5(b). The pattern-recognition step of the model would then try to find the harmonic template which contains the best matching harmonics to these three frequency components. In the present case, the harmonic template with a fundamental frequency $F0 = 200$ Hz gives the best match (Table 1.1).

If the carrier frequency f_c is now slightly changed to 2030 Hz, one obtains the waveform shown in Fig. 1.5(c). The perceived pitch is modified to a most prominent



(a) Waveform: $f_c = 2000$ Hz, $f_m = 200$ Hz (b) Spectrum: $f_c = 2000$ Hz, $f_m = 200$ Hz



(c) Waveform: $f_c = 2030$ Hz, $f_m = 200$ Hz (d) Spectrum: $f_c = 2030$ Hz, $f_m = 200$ Hz

Figure 1.5: Temporal waveform and frequency spectrum of two amplitude modulated sinusoids with carrier frequency f_c and modulation frequency f_m. (a)-(b) $f_c = 2000$ Hz and $f_m = 200$ Hz. (c)-(d) $f_c = 2030$ Hz and $f_m = 200$ Hz.

pitch frequency of 203 Hz, with some people also perceiving an ambiguous pitch at 226 Hz or 184 Hz. Can a pattern-matching paradigm explain this pitch change? A frequency analysis would retrieve three components at 1830, 2030, and 2230 Hz, as shown in the spectrum in Fig. 1.5(d). From Table 1.1, we can see that harmonics 9-10-11 of a fundamental frequency equal to 203 Hz give a very close match to these components, which would explain the perceived pitch at 203 Hz. The two other ambiguous pitches at 226 and 184 Hz might also be explained by closely matching harmonics for F0 = 225.6 Hz and F0 = 184.5 Hz, respectively.

A few important examples of pattern-matching models of pitch perception emerged in the 1970's. The model of Goldstein (1973) uses a method similar to the simplified process presented in the previous examples. After a spectral analysis in both ears,

F0	Harmonic number										
1	2	3	4	5	6	7	8	9	10	11	12
200	400	600	800	1000	1200	1400	1600	***1800***	***2000***	***2200***	2400
203	406	609	812	1015	1218	1421	1624	***1827***	***2030***	***2233***	2436
225.6	451	677	902	1128	1354	1579	***1805***	***2030***	***2256***	2482	2707
184.5	369	554	738	923	1107	1292	1476	1661	***1845***	***2030***	***2214***

Table 1.1: Frequencies in Hz of the twelve first harmonics for four different fundamental frequencies. Best matching harmonics for two tone complexes at $(1800, 2000, 2200)$ Hz and $(1830, 2030, 2230)$ Hz are emphasized.

Gaussian noise is added to each independent channel. A central processor then assesses the fundamental frequency of the stimulus with a statistical estimation of harmonic sets giving the best match.

Terhardt (1972) proposed a similar model. He assumed that the pitch was always a subharmonic of a resolved partial. After frequency analysis, his model thus first calculates successive subharmonics for all partials of the stimulus. The subharmonic with the highest coincidence across channels is then used as the cue to pitch. Let us consider the previous example with the stimuli from Fig. 1.5 again. With the 2000-Hz carrier, all three frequency components $(1800, 2000, 2200$ Hz$)$ have a subharmonic at 200 Hz (Table 1.2). This high coincidence is used to determine the produced pitch at 200 Hz. In the same fashion, with the 2030-Hz carrier, all three partials $(1830, 2030, 2230$ Hz$)$ have a subharmonic around 203 Hz, which corresponds to the most prominent pitch. One can see from Table 1.2 that the coincidence is also high around 226 and 185 Hz, which could explain the two ambiguous pitches heard by some listeners.

f	Subharmonic number										
1	2	3	4	5	6	7	8	9	10	11	12
1800	900	600	450	360	300	257	225	***200***	180	164	150
2000	1000	667	500	400	333	286	250	222	***200***	182	167
2200	1100	733	550	440	367	314	275	244	220	***200***	183
1830	915	610	458	366	305	261	*229*	***203***	*183*	166	153
2030	1015	667	508	406	338	290	254	*226*	***203***	*185*	169
2230	1115	743	558	446	372	319	279	248	*223*	***203***	*186*

Table 1.2: Frequencies in Hz of the twelve first subharmonics for six different frequencies. Subharmonics of highest coincidence for two tone complexes at $(1800, 2000, 2000)$ Hz and $(1830, 2030, 2230)$ Hz are emphasized.

Harmonic templates

Pattern-matching models all assume that sets of harmonic or subharmonic templates preexist in our memory, and that we are able to compare the frequency pattern derived from a given stimulus to these templates. This assumption requires that we have learned (sub)harmonic templates beforehand. Terhardt (1974) suggested that we learn to associate a single frequency component to its subharmonics through early exposure to harmonic-rich sounds (*e.g.*, speech and music). According to this, he added a learning matrix to his pitch perception model. However, studies performed on infants later revealed that they had an innate sense of pitch (Clarkson and Rogers, 1995; Montgomery and Clarkson, 1997). This would mean that, if there is a learning process of harmonic templates, it has to take place before birth, but no harmonic-rich sounds are perceivable from inside the womb.

The necessity of learning of harmonic templates, combined with a lack of supportive biological evidence, has given rise to some criticism of pattern-matching models of pitch perception. However, Shamma and Klein (2000) showed that harmonic templates could emerge naturally from input stimuli without a harmonic structure (*e.g.*, broadband noise or click trains). Using a simple model of the peripheral auditory system with its output fed to a coincidence matrix, they found that regions of high coincidence in the matrix output were exactly spaced at harmonic distances. This would mean that harmonic templates are not necessarily learned through exposure to a specific stimulus type, but are a natural consequence of the basic processing properties in the early stages of the auditory system. According to Shamma and Klein (2000), the templates emerge due to the nonlinearities in peripheral auditory processing. In that respect, the crucial stages in their model are the nonlinear transformations such as half-wave rectification and temporal sharpening, the sharp frequency resolution of the cochlea, as well as the rapid phase shifts of the traveling wave on the basilar membrane near the resonance point. These findings might be an important step favoring a new generation of pattern-matching models, as they also bring more biological plausibility.

1.3.2 Time theory

Introduction

According to the time theory of pitch perception, the precise temporal coding in auditory-nerve fibers is exclusively used by the system to extract the pitch of sounds. As illustrated in Fig. 1.6, the input sound generates different patterns of vibration at different places on the basilar membrane. These vibrations then create synchronous firing of auditory-nerve fibers. At frequencies up to about 5 kHz, this is called *phase locking*, as the fibers fire exactly when the basilar membrane waveform has a certain phase value. Time intervals between successive firings, also called inter-spike intervals (ISI), occur approximately at integer multiples of the waveform period. It is these intervals that are then assumed to be used by the system to determine pitch.

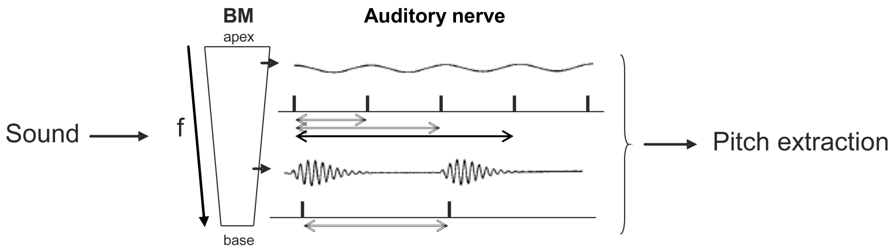


Figure 1.6: Simplified illustration of the time theory of pitch perception. f: frequency. BM: Basilar Membrane.

Historical background

In the times of Ohm (1843) and Helmholtz (1877) and the dominance of their place theory, a more modest concurrent idea was raised by Rutherford (1886), who claimed that the ear simply transmitted sounds to the brain in the same way as a telephone receiver. The so-called “telephone theory” was indeed based on the principle of a continuous waveform converted to short impulses with a precise timing, as we find in auditory-nerve fibers. The firing rate would then be the cue to pitch. However, the question remained at the time whether auditory nerve fibers could fire fast enough to account for high pitches. In their “volley theory”, Wever and Bray (1930) suggested that, as there were multiple fibers, these could fire in turn so that they could, together, produce high firing rates. Unfortunately, measurements later proved that the firing was stochastic and limited to about 300 spikes per second. Despite this, it was eventually shown that time structure could actually be carried by the pattern of instantaneous firing probability up to about 5 kHz (Johnson, 1980).

Temporal models

At first, it was often believed that the pitch of complex stimuli followed their temporal envelope. But we already saw that two stimuli with the same temporal envelope could elicit two different pitches (*cf.* Fig. 1.5(a) and 1.5(c)). de Boer (1956a) and Schouten *et al.* (1962) also showed that the pitch was shifted if partials of a modulated carrier were mistuned by equal amounts. Therefore, envelope is not a satisfactory cue to pitch, and most modern temporal models of pitch perception use a much more powerful tool: autocorrelation.

A neuronal autocorrelator was first proposed by Licklider (1951). In his model, a cochlear frequency analysis is first performed. Then, in each independent frequency channel i, corresponding to place x_i on the basilar membrane, inner-hair cells send their signal to a group of input-neurons A. These input-neurons forward the original signal to a chain of delay-neurons B_k. Each group of neurons B_k then sends a delayed version of the original signal to neurons C_k, with a delay of k times a reference synaptic delay. C_k neurons act as autocorrelators. They receive both the original non-delayed signal from A neurons and the delayed signal from B_k neurons. C_k neurons will fire if neurons A and B_k fire silmutaneously, *i.e.*, if the delayed and original signals are identical. This corresponds to an autocorrelation operation (Eq. 1.3), for which the output is maximal when the delayed and non-delayed inputs are identical. A measure of the excitation of neurons D_k, that receive the output from C_k, is then used to derive a two-dimensional pattern of excitation, function of the delay τ and the channel frequency x. A ridge along the frequency dimension x appears due to maximal excitation of D neurons at delays corresponding to the stimulus period. The interval between these ridges is used to estimate the period T, the pitch corresponding to its inverse $1/T$.

In a computer model inspired by that of Licklider (1951), Meddis and Hewitt (1991) implemented more realistic filtering and transduction stages. In particular, they took outer and middle-ear filtering, hair-cell transduction, and the refractory period of auditory-nerve fibers into account. Their model also derives a two-dimensional pattern, function of channel frequency and delay, in the form of an ISI autocorrelogram. An additional step in their model calculates a summary autocorrelation function (ACF), by simply summing the model output accross frequency. The interval between two maxima in the summary ACF gives an estimation of the stimulus period, hence the pitch.

Other operations than autocorrelation, though related to it, have been suggested in various temporal models to measure self-similarity. One of them is the difference function (Eq. 1.4), that has been used by de Cheveigné (1998) in his "cancellation model" of pitch perception. The difference function d is related to the signal energy e and the autocorrelation function r as follows:

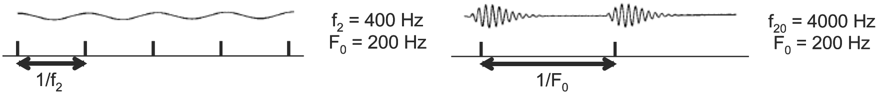
$$\begin{aligned} d(\tau) &= \frac{1}{2}\int [x(t) - x(t-\tau)]^2 dt \\ &= \int x(t)^2 dt - \int x(t)x(t-\tau)dt \\ &= e - r(\tau) \end{aligned} \tag{1.4}$$

When the delayed and non-delayed signals are identical, *i.e.*, for delays equal to multiples of the stimulus period, the autocorrelation function shows a maximum and the difference function a minimum at zero. Although they are related to each other, these two operations require different mechanisms when performed by arrays of neurons: While autocorrelation relies on excitation, cancellation relies on inhibition, and it seems physiologically reasonable to assume that not all synaptic mechanisms are excitatory.

Resolved and unresolved harmonics

Performing autocorrelation, or a similar operation, on the temporal pattern of auditory-nerve firing corresponds to searching for the best coincidence among ISIs. It is in that sense important to notice that ISIs do not give us the same information at low *vs.* high frequencies:

- Lower harmonics of a complex stimulus are *resolved*: Each harmonic falls into a single auditory filter and produces a local peak of excitation on the tonotopic axis. The ISIs in corresponding auditory-nerve fibers are thus related to the frequency of each individual harmonic (Fig. 1.7(a));
- Higher harmonics of a complex stimulus are *unresolved*: Several harmonics fall into the same auditory filter and the basilar membrane vibration pattern is due to the interference between these harmonics. Therefore, ISIs in corresponding auditory-nerve fibers are related to the overall period of the whole stimulus (Fig. 1.7(b)).



(a) Resolved harmonics (b) Unresolved harmonics

Figure 1.7: Relationship between harmonic resolvability and inter-spike intervals (ISIs) in auditory-nerve fibers. Example of basilar membrane waveform and ISI for two harmonics with a fundamental frequency F0 = 200 Hz. (a) Second harmonic $f_2 = 400$ Hz: Lower harmonics are resolved and ISIs are multiples of the period of the individual harmonic $1/f_2$. (b) Twentieth harmonic $f_{20} = 4000$ Hz: Higher harmonics are unresolved and ISIs are multiples of the overall period of the stimulus $1/\mathrm{F0}$.

Pattern-matching models of pitch perception (described in section 1.3.1) require harmonics to be resolvable to extract the pitch of complex stimuli, whereas temporal models can estimate the stimulus period from the interaction between unresolved harmonics. This is a strong advantage of temporal models over purely spectral models of pitch perception: Psychophysical experiments have indeed shown that a complex consisting only of high unresolvable harmonics could still elicit a pitch at the fundamental frequency (Ritsma, 1962). Unresolved harmonics must thus interact to allow pitch extraction. Houtsma and Goldstein (1972) later found that two harmonics presented dichotically (*i.e.*, one in each ear) could also elicit a pitch at F0. This would mean that interaction in the cochlea is not required and that, in such a case, harmonics must interact centrally.

The question of harmonic resolvability might so far seem to favor temporal models of pitch perception. However, several studies have suggested that different pitch mechanisms could be involved for resolved *vs.* unresolved harmonics. For instance, Houtsma and Smurzynski (1990) found that F0 discrimination was much worse when the pitch was retrieved from unresolved harmonics than when resolved harmonics were used, even when harmonics lied in the same frequency region. A study by Plack and Carlyon (1995) also showed that F0 discrimination improved faster with tone duration for unresolved than resolved harmonics. Different integration mechanisms could thus be involved. The fact that purely temporal models of pitch perception use the same algorithm to extract the pitch from resolved and unresolved harmonics is not consistent with these findings, and this sets a new limit to their advantage over purely spectral models.

1.4 Place *vs.* Time

Modern models of pitch perception based on either place or time theories account for pitch perception of a wide range of common natural stimuli equally well. Therefore, most recent studies have to use very specific, more artificial, stimuli in order to see where the weak points of each model lie. We have already mentioned some limitations of the two main types of pitch perception models:

- *Spectral models*, most of which are based on pattern-matching schemes, require a set of preexisting harmonic templates, which raises the question of learning, although these templates might emerge naturally because of the nonlinear properties of the peripheral auditory system. These models also require some partials of the stimulus to be resolved for pitch extraction, and even though most natural sounds contain resolved components, a low periodicity pitch can still be produced by unresolved harmonics.

- *Temporal models*, most of which are based on autocorrelation schemes, require precise phase-locking in auditory-nerve fibers, and this only works below about 5 kHz. However, this might not be too critical as most natural sounds have a fundamental frequency below 5 kHz, and pitch is sometimes not considered musical above that frequency. In spite of this, another limitation comes from the fact that temporal models use the same algorithm to extract pitch from resolved and unresolved partials, which is not necessarily justified.

Both theories and their corresponding models seem to have their advantages and drawbacks. Which direction shall we then take? Is there a way to test whether one theory can be validated by finding evidence against the other one? Or is this division into two concurrent theories misleading? In that case, it might be more sensible to develop a new approach in the light of current knowledge.

A first idea could be to ask why the auditory system couldn't use both cues for pitch extraction, as it is able to code both place and time precisely. For example, the sluggishness of phase-locking at high frequencies may suggest that time cues are used below about 5 kHz, and place cues above that frequency. However, one may as well justify the somewhat opposite suggestion that place cues are used at low frequencies because of resolvable harmonics and time cues at high frequencies because of unresolvable harmonics. It is thus still unclear whether these frontiers, defined by the phase-locked region and the region of harmonic resolvability, are relevant for pitch-related mechanisms.

In some recent combined models, both place and time play a role in the auditory processing leading to pitch determination: In a model by Moore (2003), ISIs are analyzed separately in each channel, after what a search for common intervals is run across characteristic-frequency channels. The most prominent interval at the output, generally that corresponding to the fundamental frequency, determines the pitch.

Other models assume spectrotemporal neural mechanisms comparing the relative timing of spikes in neighboring auditory channels. This may be achieved *via* lateral inhibitory processes (*e.g.*, Shamma, 1985) or *via* coincidence detection across channels (*e.g.*, Carney, 1994). Using the former approach, Cedolin and Delgutte (2010) argued that such a use of direct spectrotemporal representations at the level of the auditory nerve was more consistent with current psychophysical knowledge than the classical rate-place and ISI pitch representations, as it could account for the relative independence of pitch salience on level, as well as its degradation with increasing frequency and decreasing harmonic resolvability.

Despite such attempts to model the use of both spectral and temporal information for pitch extraction, the debate of "place *vs.* time" is still ongoing in pitch perception research. Following the findings of Plomp (1967) and Ritsma (1967), who showed that the low, resolved harmonics dominated the pitch percept, and the discovery that cochlear distortion products also influenced complex pitch (Smoorenburg, 1970), opinions tended to favor the place theory in the 1970's. With the advancement of measurement techniques, the interest in temporal pitch codes based on synchrony was then revived, as physiological evidence for accurate representations of phase-locked information flourished (*e.g.*, Johnson, 1980). The time theory has thus regained devoted supporters, also because some pitch phenomena cannot be accounted for by the place theory: For instance, it is possible to match a sinusoid to a noise signal abruptly turned on and off periodically, although no spectral cues are available (Pollack, 1969). The same is true for binaural-pitch stimuli, that can be created by interaural phase delays in white noise (*e.g.*, Cramer and Huggins, 1958), and contain no spectral information *per se*. Moreover, temporal cues are widely assumed to be used extensively by the binaural system, and it is tempting to think that some physiological units used by the binaural system might also be involved in the processing of pitch-evoking stimuli. Finally, recent studies by Moore *et al.* (2009b) and Moore and Sęk (2009a) suggested that the information contained in the fast temporal fluctuations, or temporal fine structure, of internal sound representations was responsible for discrimination of harmonic and frequency-shifted complex tones using a pitch cue, even in spectral regions above 5 kHz.

In spite of this regained interest in theories and models favoring timing rather than place, a few recent studies have focused on the comparison of time and place mechanisms by using specific stimuli. In particular, Oxenham *et al.* (2004) investigated the perception of the pitch of so-called “transposed stimuli” (van de Par and Kohlrausch, 1997). The purpose of such stimuli is to present a low-frequency temporal fine structure to high-frequency regions of the cochlea. This way, it is expected that transposed stimuli should elicit a low-frequency pitch if the system relies on temporal information, and a high-frequency pitch or no pitch if the system relies on spectral information. A way to generate a transposed tone is to multiply a high-frequency carrier with a half-wave rectified sinusoid. After half-wave rectification and low-pass filtering by the hair-cell transduction process, the transposed tone is assumed to have the same internal representation as a pure tone. Oxenham *et al.* (2004) performed F0 discrimination and pitch-matching experiments using both regular and transposed single and complex tones. Their results showed that performance in F0 discrimination was poorer with transposed tones than regular tones. Moreover, their subjects could not precisely match the fundamental pitch of a transposed complex tone consisting of three transposed harmonics to a regular tone. Because their subjects could not extract the fundamental frequency from a group of transposed harmonics, they concluded that it was necessary to present harmonics at the right place in the cochlea in order to perceive a pitch at F0. As argued by Shamma (2004), the time theory would predict a clear pitch at F0 for such stimuli, whereas the place theory predicts no such pitch. This would mean than tonotopicity has to be preserved to extract the fundamental pitch of complex tones, and that if temporal cues are used, they cannot do without the presence of accurate place cues.

Such studies have revived the debate about the use of place *vs.* time cues by the auditory system for pitch extraction. As mentioned above, approaches that do neither refer to purely temporal nor to purely spectral cues, but assume the use of combined spectrotemporal information by the auditory system, also offer plausible alternatives. It is after all known that many auditory neural units, in addition to their ability to convey information that is precise in time, are tuned to characteristic frequencies for which their probability of firing has a maximum. Neurons are thus able to convey both spectral and temporal information. The crucial question now remains of how this information is coded by neural units contributing to pitch extraction.

1.5 Summary

The question of how the human auditory system extracts the pitch of complex sounds has historically been addressed *via* two concurrent approaches:

- The place theory, according to which spectral cues are essential for pitch perception, has given rise to a number of models relying on pattern-matching schemes: Tonotopic frequency information is used to derive a pattern that is then matched to preexisting harmonic templates. These templates are either learned and then retrieved from memory, or may originate naturally from the properties of the peripheral auditory system itself. The fact that such models only make use of the resolvable components in the stimulus, together with the lack of biological evidence towards the mechanisms they involve, has put the dominance of the place theory to the side at the end of the 20th century, in favor of the time theory of pitch perception.
- The time theory claims that the timing of neural spikes from the auditory nerve and onwards provides the key information to central units for pitch extraction. A large variety of models have exploited this assumption, using operations that are often similar to autocorrelation. Temporal models have the advantage that they can use information from unresolvable components as well as resolvable components in the stimulus. However, they do not take resolvability into account for how the different components are processed, although this aspect might influence perception. Another source of criticism comes from the fact that purely temporal models do not seem to predict the pitch of transposed stimuli correctly. This means that accurate spectral information may still be required for temporal cues to be useable by the system.

Despite their fundamental differences, both spectral and temporal models can accurately predict the pitch of most pitch-evoking stimuli. In order to understand the mechanisms underlying the pitch percept, one must therefore use pitch-evoking stimuli for which existing theories and models do not bring a satisfactory explanation. Binaural pitch and transposed tones are good examples of such stimuli, and were therefore chosen as reference stimuli in the present work. Studying the effects of hearing impairment on pitch perception is also essential: If we can identify which specific auditory functions make the pitch percept fail, we might more easily determine which mechanisms are more likely to be involved in pitch perception. For that purpose, it might be useful to keep a spectrotemporal approach in mind, as both spectral and temporal cues are likely to play a role, not only individually at different levels of the auditory pathway, but maybe also in more complex combined mechanisms.

1.6 Outline of the present work

This first chapter aimed at introducing some basic definitions and a brief overview of the history and advancements of pitch perception research. The following chapters will focus on reporting and discussing the findings of the present work. The aim was to better understand the nature of the key mechanisms involved in the internal representation of pitch, by performing psychophysical experiments in normal-hearing and hearing-impaired listeners, using stimuli with specific spectrotemporal properties. The role of spectral *vs.* temporal information for pitch perception, as well as the correlation between pitch perception outcomes and measures of specific auditory functions, were the key aspects addressed here. The thesis can be divided into two main parts:

The first part consists of chapters 2 to 4, which are concerned with the perception of binaural-pitch stimuli in normal-hearing, hearing-impaired, and reading-disabled listeners:

- In **Chapter 2**, the perception of binaural-pitch stimuli in normal-hearing and hearing-impaired listeners is investigated. The effects of hearing impairment on binaural-pitch detection, binaural-pitch discrimination, and melody recognition with different binaural-pitch types are studied. The results of individual listeners are compared to measures of spectral and temporal resolution, in an attempt to determine which auditory functions are crucial for binaural pitch perception.

- In **Chapter 3**, the presence of a link between impaired binaural pitch perception and a specific auditory deficit is investigated in more detail, in order to further evaluate the potential of binaural-pitch stimuli as useful tools for clinical diagnoses. For this purpose, a test battery is designed to characterize each listener's individual "auditory profile", including measures of loudness perception, cognitive ability, binaural processing, temporal fine structure processing, and frequency selectivity. Besides the potential correlation of such measures with binaural-pitch perception outcomes, it is also aimed to develop relevant measures for hearing-loss characterization, *via* a study of the relationship between the different auditory-profile outcomes.
- In **Chapter 4**, it is investigated whether binaural pitch perception is impaired in listeners with developmental dyslexia. Following contradictory outcomes in earlier studies of binaural processing in dyslexic listeners, it is clarified here whether the perception of binaural-pitch stimuli differs between adults with dyslexia and matched controls.

The second part consists of chapters 5 to 7, which are concerned with pitch perception of high-frequency complex tones in normal-hearing listeners:

- In **Chapter 5**, the perception of the low pitch of high-frequency transposed tones is investigated in normal-hearing listeners. Using inharmonic stimuli with assumed unresolved components, the hypotheses of a pitch relying on temporal envelope cues *vs.* temporal fine structure cues are tested in a pitch-matching experiment. The resolvability of the stimulus components by the listeners is also evaluated. The aim is to assess whether temporal fine structure cues persist at high frequencies and whether temporal envelope cues take over as the center frequency of the complex is raised or the component spacing is lowered.

- In **Chapter 6**, it is attempted to establish a relationship between the pitch matches obtained in Chapter 5 and the performance of the listeners in discriminating harmonic and frequency-shifted high-frequency complex tones. In order to evaluate the audibility of combination tones in the pitch-matching and discrimination experiments, the level of the most prominent combination tone is measured for the stimuli of interest, as well as the background-noise level required for adequate masking. The effects of combination-tone audibility on the perceived pitch are then directly measured in a pitch-matching experiment. The influence of relative component phases and dichotic presentation on the low pitch are also investigated, and the resolvability of the components is further evaluated, with the aim to determine whether temporal or spectral cues are used for pitch extraction.

- In **Chapter 7**, the implications of the findings of Chapter 6 for the underlying pitch mechanisms are discussed. In order to assess the nature of such mechanisms, internal spatiotemporal activity patterns are obtained by feeding the stimuli to a peripheral auditory model, and pitch predictions are made based on three classical theories of pitch perception (spectral, temporal, and spectrotemporal).

Finally, the main findings of the present work are summarized, and their possible implications discussed, in **Chapter 8**.

2

Binaural pitch perception in normal-hearing and hearing-impaired listeners*

The effects of hearing impairment on the perception of binaural pitch stimuli were investigated. Several experiments were performed with normal-hearing and hearing-impaired listeners, including detection and discrimination of binaural pitch, and melody recognition using different types of binaural pitches. For the normal-hearing listeners, all types of binaural pitches could be perceived immediately and were musical. The hearing-impaired listeners could be divided into three groups based on their results: a) some perceived all types of binaural pitches, but with decreased salience or musicality compared to normal-hearing listeners; b) some could only perceive the strongest pitch types; c) some were unable to perceive any binaural pitch at all. The performance of the listeners was not correlated with audibility. Additional experiments investigated the correlation between performance in binaural pitch perception and performance in measures of spectral and temporal resolution. Reduced frequency discrimination appeared to be linked to poorer melody recognition skills. Reduced frequency selectivity was also found to impede the perception of binaural pitch stimuli. Overall, binaural pitch stimuli might be very useful tools within clinical diagnostics for detecting specific deficiencies in the auditory system.

---

* This chapter is based on Santurette and Dau (2007).

2.1 Introduction

Pitch is a very important auditory sensation that plays a major role in auditory object formation and speech perception. It is also what gives us a musical perception of sounds. Because pitch is a subjective attribute, understanding and modeling pitch perception has been a constant challenge for auditory research in the past decades. The phenomenon of binaural pitch, also called dichotic pitch, was first studied in detail by Cramer and Huggins (1958). Binaural pitch can be produced by presenting a random noise stimulus to one ear, while a slightly different random noise stimulus is simultaneously presented to the other ear. The difference introduced between the ears is most commonly the phase shift of some specific frequency components in either the left or the right stimulus. Because of the introduction of this particular interaural phase-difference spectrum, the combination of the left and right noises creates an additional auditory object in the form of a pitch sensation around a particular frequency, *i.e.*, the listener is able to hear a more or less distinct tone in the noise, whereas only noise can be perceived when listening to each stimulus independently in one ear. The visual phenomenon of stereopsis (Julesz, 1971), produced by the presentation of two slightly different random sets of dots to each eye, thus giving rise to the perception of a three dimensional image, is often used as a simple visual analogy to what binaural pitch is to audition.

Several studies have investigated different aspects of binaural pitch perception in normal-hearing listeners, using and comparing different pitch types, such as the Huggins' pitch (Cramer and Huggins, 1958), the binaural edge pitch (Klein and Hartmann, 1981), and the binaural coherence edge pitch (Hartmann, 1984b), further described in the following. Akeroyd *et al.* (2001) showed that, in normal-hearing listeners, the perception of these three pitch types was immediate, and compared their salience using a melody recognition task. Several attempts have been made to account quantitatively for binaural pitch perception, and several models have been suggested for the processing of binaural pitch. The central activity pattern (CAP) model (Raatgever and Bilsen, 1986) and models based on the equalization-cancellation (EC) model by Durlach (1960) all use temporal cues to process binaural stimuli and account for their perception. While the CAP model uses a tuning of the signals both in frequency and interaural time difference in the peripheral channels to derive a two-dimensional central activity pattern, EC models assume that the binaural system first equalizes the signals in corresponding channels, using amplitude adjustment and phase shifting, before subtracting them to obtain a residual activity spectrum (*e.g.* Durlach, 1960; Culling *et al.*, 1998). Comparing results from psychophysical experiments to model predictions, Hartmann and Zhang (2003) found that the CAP model did not predict the salience of different versions of the Huggins' pitch and the binaural coherence edge pitch correctly, while the results favored an EC process. Experiments carried out by Culling *et al.* (1998) also led to the conclusion that, although the CAP model was able to accurately predict the lateralization of binaural pitch, it did not correctly predict the range of transition bandwidths over which the Huggins' pitch and binaural edge pitch are audible, contrary to their modified version of the EC model (mEC).

The aim of the present study was to investigate the effects of hearing impairment on binaural pitch perception. Because binaural pitch signals are easy to generate and their perception seems to be immediate (Akeroyd *et al.*, 2001), it would be of interest to find a link between the ability of hearing-impaired listeners to perceive such signals and the deficiency of a particular auditory function. If this can be achieved, binaural pitch signals could become useful test-stimuli for specific clinical diagnostics. In order to identify and understand the mechanisms underlying binaural pitch perception and how these mechanisms are affected by hearing impairment, some fundamental questions first need to be addressed: Are hearing-impaired listeners able to perceive binaural pitch signals at all? If yes, is that perception immediate as in normal-hearing listeners? Are different binaural pitch types perceived and affected in the same way in a hearing-impaired listener? Finally, in the case that hearing-impairment affects binaural pitch perception, are hearing-threshold elevation, reduced frequency selectivity, and reduced temporal resolution among the key factors in the impaired listener's difficulties with binaural pitch? These fundamental aspects were the main concerns of the present study.

The influence of hearing impairment on binaural pitch perception was investigated in a set of psychophysical experiments including a melody recognition test similar to that performed by Akeroyd *et al.* (2001). The detection and discrimination of Huggins' pitch as a function of frequency was also tested, in experiments similar to those of Hartmann and Zhang (2003). Hearing-impaired listeners participated in additional experiments evaluating several auditory abilities with which patients with a hearing loss are commonly found to have difficulties: frequency discrimination, frequency selectivity, and temporal resolution. The purpose was to investigate if a link could be made between the subjects' performance in each of these tasks and their performance in the binaural pitch experiments. Frequency discrimination abilities are expected to be directly correlated to the performance of the subjects in pitch discrimination and melody recognition tasks, because these experiments require the listener to have a musical perception of sounds, and thus to be able to discriminate sounds producing close pitch frequencies. An impaired frequency selectivity, associated with a broadening of the auditory filters at the level of the basilar membrane, can also give rise to changes in the internal representation of the interaural correlation spectrum of binaural signals. Therefore, a listener with a sensorineural hearing loss involving a deficit in frequency selectivity might be expected to have more difficulties perceiving binaural pitch than a normal-hearing listener. Furthermore, as binaural pitch perception involves an analysis and comparison of the temporal fine structure of the left and right signals by the binaural system, one may expect that deficits in temporal resolution might affect binaural pitch perception. However, because binaural pitch is a low-frequency phenomenon and becomes very difficult to perceive for pitch frequencies above about 1600 Hz (Cramer and Huggins, 1958), one can expect an elevation of the listener's hearing threshold at high frequencies not to be responsible for potential difficulties in perceiving binaural pitch.

In the following, results from three experiments involving binaural pitch will be presented first: they include Huggins' pitch detection, Huggins' pitch discrimination, and melody recognition with different types of binaural pitch stimuli. The relation of these results with measures of spectral and temporal resolution will then be investigated. Implications of the results will finally be discussed.

2.2 Method

2.2.1 Subjects

14 normal-hearing (NH) and 10 hearing-impaired (HI) subjects participated in the experiments[1]. A pure-tone audiogram was obtained for each of the subjects prior to the experiments. All NH subjects' hearing thresholds were found to be below 20 dB HL at all test frequencies, while the HI subjects obtained the various audiogram shapes shown in Figure 2.1. NH subjects were aged between 22 and 40 years. HI subjects reported different signs of hearing impairment, and their age was between 25 and 68 years. The origins of the HI subjects' hearing losses were diverse, and the different categories of impairment encountered are briefly presented in the following:

- *Subjects 1 and 2* showed audiograms that were similar to those of NH listeners. However, both subjects reported that they had problems understanding speech in noisy environments or situations with several concurrent speakers.
- *Subjects 3, 4, and 5* reported that they had been exposed to loud sounds, which had damaged some of their cochlear hair cells. Those subjects all showed a high-frequency hearing loss, especially around the resonance frequency of the ear canal (3-4 kHz), but had normal hearing at low frequencies. Subject 5 had the most severe hearing loss while subject 3 showed a mild impairment. All three subjects could understand speech very well in quiet environments, but indicated that they had difficulties with speech intelligibility in noisy environments.
- *Subject 6* had normal hearing at low frequencies, and started suffering from hearing problems progressively after the age of 60, which suggests the sensorineural hearing loss was age-induced.
- *Subject 7* suffered from Ménière's disease. Symptoms involved in Ménière's disease typically include vertigo and a low-frequency hearing loss, which can

---

[1] All experiments were approved by the ethics committee of Copenhagen county (Den Videnskabsetiske Komite for Københavns Amt).

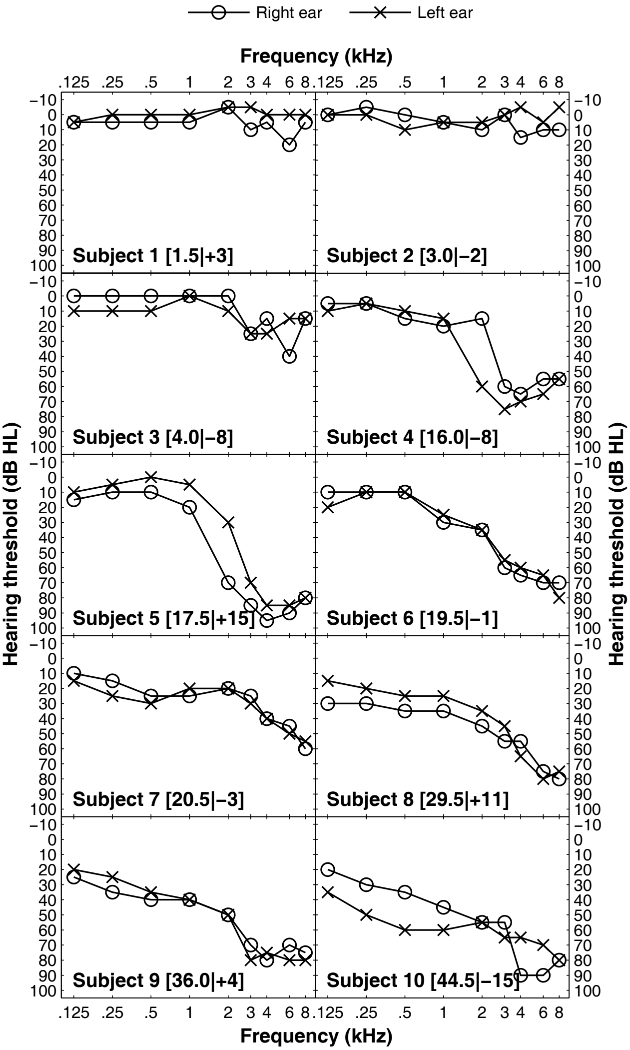


Figure 2.1: Audiograms of the 10 hearing-impaired subjects. In the bottom-left corner of each audiogram are also indicated: [the average hearing loss of the subject in dB HL between 125 and 2000 Hz|an asymmetry factor corresponding to the average difference between the right and left hearing thresholds in dB HL between 125 and 2000 Hz].

vary with time, combined with tinnitus and sensations of fullness in the ear. Subject 7's audiogram showed only a mild hearing loss at low frequencies, and a progressively more severe hearing loss at frequencies above 2 kHz.

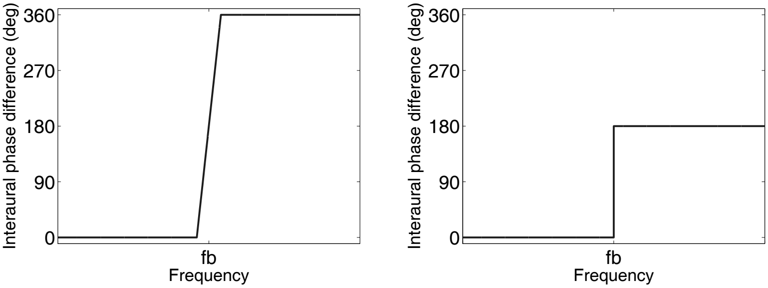
- *Subjects 8, 9, and 10* had sensorineural hearing losses, with audiograms showing a slope falling down towards high frequencies. *Subject 8* had been operated following an otosclerosis infection. *Subject 9* had been suffering from diabetes for her whole life, and suggested her hearing problem was probably linked to the disease, which had progressively affected other perceptual abilities like her eyesight. The hearing loss of *subject 10* was a consequence of cranial trauma following a car accident, which had affected other perceptual abilities like olfaction, which suggests the hearing loss may be due to damage of central areas of the auditory system.

2.2.2 Stimuli

Six binaural pitch types with different frequency-dependent interaural phase-difference patterns were used in the experiments, indicated in Figure 2.2.

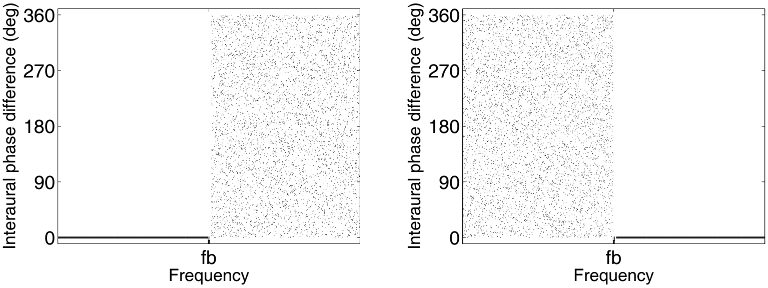
In the *Huggins' pitch* (HP) phase-difference pattern (Cramer and Huggins, 1958), the signals in both channels are in phase at all frequencies except a narrow frequency range around the boundary frequency, f_b (Figure 2.2(a)). In the transition area around f_b, a linear phase-shift is introduced such that, if the transition bandwidth is b, the phase-difference between both channels varies linearly from 0 to 2π in the frequency interval $[f_b - b/2, f_b + b/2]$. The phase-difference is maximum (π) at the boundary frequency. Detection of HP is optimal for transition bandwidths between 4% and 32% of the boundary frequency (Culling *et al.*, 1998).

In the *binaural edge pitch* (BEP) phase-difference pattern (Klein and Hartmann, 1981), the signals in both channels are in phase below the boundary frequency (or edge frequency) f_b, and in antiphase above f_b (Figure 2.2(b)). In the BEP configuration, no transition is needed for optimal detection of the pitch, but a linear phase shift from 0 to π may as well be used without altering the pitch detection, provided the transition bandwidth is below 8% of the boundary frequency (Culling *et al.*, 1998).



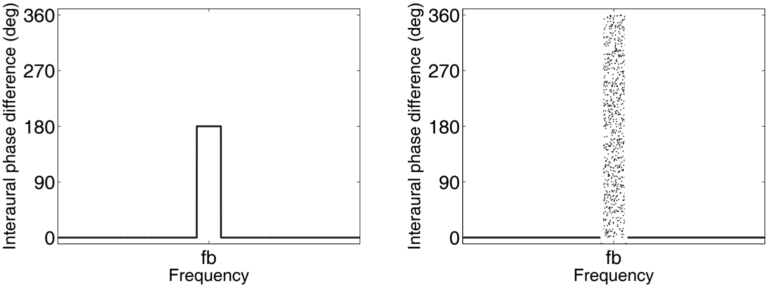
(a) Huggins' pitch (HP)

(b) Binaural edge pitch (BEP)



(c) Binaural coherence edge pitch (BICEP)

(d) Reversed BICEP (RBICEP)



(e) Binaural band pitch (BBP)

(f) Binaural incoherent-band pitch (BIBP)

Figure 2.2: Six phase-difference patterns producing a binaural pitch sensation. (a) Huggins' pitch: linear phase shift from 0 to 2π around the boundary frequency f_b. (b) Binaural edge pitch: constant phase shift of π above f_b. (c) Binaural coherence edge pitch: random phase shift above f_b. (d) Reversed binaural coherence edge pitch: random phase shift below f_b. (e) Binaural band pitch: constant phase shift of π around f_b. (f) Binaural incoherent-band pitch: random phase shift around f_b.

In the *binaural coherence edge pitch* (BICEP) phase-difference pattern (Hartmann, 1984b), the signals in both channels are in phase below the boundary frequency (or edge frequency) f_b, and a random phase difference is introduced at all frequencies above f_b (Figure 2.2(c)). In other words, the left and right noises are coherent below f_b and incoherent above f_b, thus creating a coherence edge. Reversing the coherent and incoherent parts of the noise, *i.e.*, introducing the random phase difference at all frequencies below f_b and keeping the noises in phase above f_b, produces a similar pitch sensation (Hartmann, 1984b). This configuration will be referred to as *reversed binaural coherence edge pitch* or RBICEP (Figure 2.2(d)).

The four configurations presented above (HP, BEP, BICEP, and RBICEP) are the most common patterns for generation of a binaural-pitch signal using a frequency dependent phase-difference between both channels. Two additional configurations were also investigated in the present study. One configuration is a stimulus which was used by Yost (1991) and Zhang and Hartmann (2004), and resembles the HP configuration except that the phase in a narrow bandwidth around f_b is shifted by a fixed amount in one channel. Thus, the phase difference between both channels is constant as a function of frequency, and there are two sharp edges similar to the edge of the BEP configuration. A phase shift of π will be used in this study (Figure 2.2(e)), and the configuration will be called *binaural band pitch* (BBP). The second additional configuration is similar to the BBP configuration except that the edges are coherence edges, which means that the phase in a narrow bandwidth around f_b is randomized in one channel (Figure 2.2(f)). This configuration, similar to the stimulus used in Experiment 1 of Akeroyd and Summerfield (2000), will be called *binaural incoherent-band pitch* (BIBP).

The binaural pitch signals were generated as follows: (1) Random noise with the desired duration was generated in the spectral domain, using a sampling rate of 44100 Hz. (2) The amplitude was adjusted such that the noise had a constant amplitude and only the phase was randomized, *i.e.*, white noise was created. (3) All frequency components above 4000 Hz were set to zero. (4) The stimulus obtained after step 3 was left unchanged and was transformed back to the time domain using the inverse FFT algorithm. It was then fed to the left channel. (5) Some of the phase components of the stimulus obtained after step 3 were modified in order to create the desired phase-difference pattern between the two ears. The stimulus was transformed back to the time domain using the inverse FFT algorithm. It was then fed to the right channel.

2.2.3 Apparatus

During all tests, subjects were sitting in a soundproof listening booth equipped with a PC. All stimuli were generated in MATLAB and converted to analog signals using a 16-bit D/A converter with a sampling rate of 44100 Hz. They were fed through the PC soundcard (RME DIGI96/8) which was connected to Sennheiser HD580 headphones. The level of the signals was calibrated by connecting the headphones to an artificial ear (IEC 318) which was itself linked to a B&K 2607 sound level meter. The presentation level of all stimuli was 65 dB SPL for the NH subjects and was adjusted for the HI subjects depending on their hearing thresholds. As reported by Cramer and Huggins (1958), variations in level do not seem to affect binaural-pitch perception much at comfortable presentation levels, as long as the signals are not too weak. In the present study, the levels were chosen such as to produce comfortable loudness in both ears.

2.3 Binaural pitch experiments

Before the different binaural pitch experiments were carried out, it was checked whether the subjects were able to hear binaural pitch at all in a preliminary test. Each listener was presented a 10-second signal, consisting of 10 noise intervals of 1 second duration each. A 30-ms cosine-ramp was applied to the onset and offset of each interval. The first and the last interval contained no binaural pitch, *i.e.*, diotic noise, while intervals 2 to 9 each contained a Huggins' pitch with values of f_b ranging from 523.25 Hz (C5) to 1046.50 Hz (C6), so that the musical C-scale was played. The transition bandwidth was equal to 16% of f_b. The task of the listener was to listen to the noise stimulus and tell the experimenter if something else than noise could be perceived. After listening to the stimulus only once, all 14 NH subjects and HI subjects

1 to 8 reported they could hear a tone rising progressively. This confirms that binaural pitch, if perceivable, is an immediate sensation and requires no learning in order to be perceived (Akeroyd *et al.*, 2001). When asked where the perceived pitch was located, 13 NH and 7 HI subjects clearly mentioned that the pitch was heard inside the head and lateralized towards the right ear, while the two remaining subjects heard the pitch more towards the back of the head. HI subjects 9 and 10 reported they could not hear anything but noise, even by listening to the stimuli several times. For these two subjects, the stimuli were also played at higher levels and with other pitch types, but in none of the configurations could the subjects perceive a pitch. This suggests that the ability of a hearing-impaired person to detect HP is either immediate or non-existing. It is of interest to mention that the two subjects who could not perceive binaural pitch at all (subjects 9 and 10) were those for whom it was most likely that central areas of the auditory system had been damaged. Those two subjects did not participate in the following experiments involving binaural pitch.

2.3.1 Huggins' pitch detection

This first experiment aimed at measuring the detectability of HP as a function of frequency. The experiment was similar to Experiment 1 of Hartmann and Zhang (2003), but a larger frequency range was used such that higher frequencies were also tested.

Method

The test consisted of 80 trials testing 8 different boundary frequencies, with 10 trials for each value of f_b. The tested boundary frequencies were 80, 125, 200, 315, 500, 800, 1250, and 2000 Hz. The order of the trials was randomized, and a 2-alternative forced-choice (AFC) procedure was used. In each trial, the listener was presented two noise intervals of 500 ms each, separated by a silence of 500 ms. A 30-ms cosine-ramp was applied to the onset and offset of each interval. One of the intervals contained a Huggins' pitch (with $b = 0.16f_b$) whereas the other interval contained no binaural pitch, *i.e.*, the noise was the same in both ears. The interval containing the pitch was

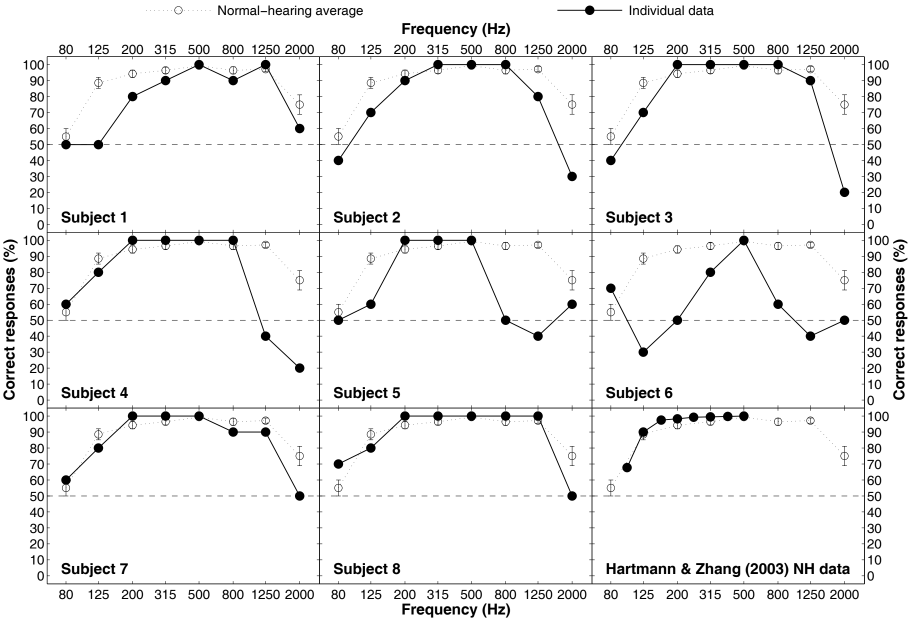


Figure 2.3: Detectability of HP for eight different boundary frequencies: results from 8 hearing-impaired subjects. The average results and standard error over 14 normal-hearing subjects are indicated in each plot by empty circles. The ninth subplot in the lower-right corner indicates the results obtained by Hartmann and Zhang (2003) with normal-hearing listeners.

random. The task of the subject was to indicate through a computer interface which interval contained the pitch. No feedback was provided.

Results and discussion

Detectability was measured as the percentage of correct responses for each boundary frequency. Chance level corresponded to 50% of correct responses. The average results over all 14 NH subjects and the individual results of the 8 HI subjects are given in Figure 2.3.

For NH listeners, the average detectability of HP appeared to be best for frequencies between 200 Hz and 1250 Hz, where it lay above 94%. The most salient frequency was 500 Hz, and this was also the frequency where the smallest standard deviation was observed. At low frequencies, the results are consistent with those obtained by Hartmann and Zhang (2003), showing a declining detectability with decreasing frequency, reaching chance level at 80 Hz. The large standard deviation at this particular frequency is due to the fact that, although most subjects scored around chance level or below, some of them could perceive the pitch even at such a low frequency. The presentation level being the same at all frequencies, the lower detectability of HP at low boundary frequencies may be due to the higher hearing threshold at these frequencies, as suggested by Cramer and Huggins (1958). Culling (1999) suggested that the use of a transition bandwidth b proportional to the bandwidth of the auditory filter centered on f_b could be better than using a fixed percentage of f_b. In this experiment, the fixed value of $b = 16\%$ of f_b appears to be close to the equivalent rectangular bandwidth (ERB) as defined by Glasberg and Moore (1990), except at frequencies below 200 Hz where the ERB progressively becomes broader than b. Whether this really has an influence on HP detectability at low frequencies and is actually responsible for the results obtained was not further investigated in the present study. It could be tested by measuring HP detectability with values of b equal to a fixed fraction of the ERB for each boundary frequency. At the boundary frequency of 2000 Hz, a large standard deviation was obtained. Here, 9 subjects showed a very good detectability (80-100%), while 5 of them obtained a score in the range 20-60%, reflecting that their detection of the pitch was non-existing or very weak. This suggests a high variability of the decline of the pitch at high frequencies among listeners. Interestingly, subjects who easily detected HP at 2000 Hz did not necessarily detect the pitch at 80 Hz, and *vice versa*, suggesting that performance in binaural pitch detection at low *versus* high frequencies might not be linked.

The HI subjects can be divided into two groups according to their results: (1) Subjects 2, 3, 7, and 8 obtained similar results as the NH subjects, with high scores (above 80%) in the frequency range between 200 Hz and 1250 Hz, and a declining detectability below 200 Hz and above 1250 Hz. The hearing impairment of these subjects did not seem to affect HP detectability. (2) The other HI subjects (1, 4, 5, 6) obtained high scores in a narrower mid-frequency range than NH listeners, and a detectability that declined faster outside this mid-frequency range than in the NH subjects. For subjects 5 and 6, both low and high frequencies were affected. For subject 4, only high boundary frequencies were significantly affected, with chance level reached at 1250 Hz and 2000 Hz. For subject 1, a poorer detectability than the NH average was obtained only for low values of f_b (200 Hz and below). Thus, for the subjects from the second group, detectability of HP appeared to be reduced either for low boundary frequencies only, high boundary frequencies only, or both low and high boundary frequencies. Thus, as in NH listeners, there is no link between performance in HP detection at low and high frequencies. Finally, HP detectability was not related to the shape of the audiogram or to the origin of the subject's hearing loss, with a correlation coefficient $r = 0.14$ between the average score of the subjects in the HP detection test and their average hearing threshold between 125 and 2000 Hz.

2.3.2 Huggins' pitch discrimination

The aim of this second experiment was to measure the ability of the subjects to discriminate two Huggins' pitches with close boundary frequencies (separated by 1.5 musical semitones), which was a way of testing the musicality of the pitch for the listeners. The experiment was similar to Experiment 2 of Hartmann and Zhang (2003), but a larger frequency range was used such that higher frequencies were also tested. For comparison purposes, the test was also performed with pure-tone signals.

Method

The test consisted of 2 runs: a first run with pure-tone signals, and a second run with HP stimuli. Each run consisted of 80 trials testing 8 different frequencies, with 10 trials for each frequency value. The tested frequencies were 80, 125, 200, 315, 500, 800, 1250, and 2000 Hz. The order of the trials was randomized, and a 2-AFC procedure was used. On each trial, two intervals of 500 ms each, separated by a silence of 500 ms, were presented to the listener. A 30-ms cosine-ramp was applied to the onset and offset of each interval. In the first run, pure-tone stimuli were used, and on each trial the frequency of the second pure tone was either 9% (or 1.5 musical semitones) lower than, 9% higher than, or the same as that of the first one, which was

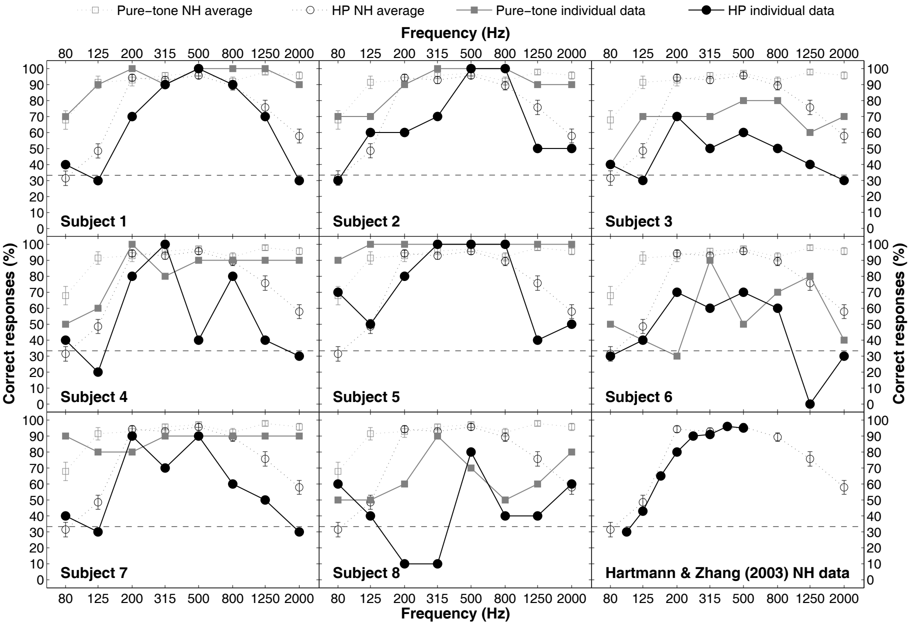


Figure 2.4: Discrimination of a 1.5-semitone interval with pure-tone and HP stimuli for eight different frequencies: average results and standard error over 14 normal-hearing subjects and results from 8 hearing-impaired subjects. The ninth subplot in the lower-right corner indicates the results obtained by Hartmann and Zhang (2003) with normal-hearing listeners.

equal to one of the values listed above. In the second run, both intervals contained a Huggins' pitch (with $b = 0.16 f_b$). The boundary frequency in the first interval was equal to one of the values listed above, while the boundary frequency in the second interval had one of the following values, chosen randomly: 9% lower than, 9% higher than, or the same value as in the first interval. The task of the subject was to indicate through a computer interface whether the pitch in the second interval was lower, the same, or higher than the pitch in the first interval. No feedback was provided.

Results and discussion

Discrimination was measured as the percentage of correct responses for each frequency. Chance level corresponded to 33% of correct responses. The average results and standard error over all 14 NH subjects and the individual results from the 8 HI subjects are given in Figure 2.4.

For NH subjects, pure tone discrimination was very good, with average scores above 91%, for all frequencies except the lowest test-frequency of 80 Hz. At such a frequency, most subjects showed a reduced ability to discriminate the tones, with scores in the range of 50-80%. One might expect that the results for HP discrimination reflect a combination of those obtained in the HP detection experiment (section 2.3.1) and the pure-tone discrimination experiment. The task required the listener to be able to detect HP and to be able to discriminate between two close pitch frequencies. In fact, at low frequencies (80, 125 Hz), the declining detectability of HP reduces HP discrimination. Discrimination appeared to be nearly impossible at 80 Hz. At 125 Hz, discrimination was very difficult or impossible for some listeners, only two subjects obtaining more than 70% correct responses. The results at low frequencies are consistent with those obtained by Hartmann and Zhang (2003) only showing a slightly higher value at 200 Hz. At high frequencies, HP discrimination becomes more difficult with increasing frequency, but the frequency at which the decay starts varies among listeners. Results in the mid-frequency range clearly show that Huggins' pitch signals are as musical as pure tones between 200 Hz and 800 Hz, and up to 1250 Hz for about half of the test subjects, the average upper-limit probably lying between 800 Hz and 1250 Hz.

Some HI subjects obtained results close to the NH average in the pure-tone condition. Only subjects 3, 6, and 8 showed a different behavior. Subject 3 had scores about 25-30% lower than the normal-hearing average at all frequencies, whereas subjects 6 and 8 showed more variable results with frequency and obtained the best discrimination at a frequency of 315 Hz. In the HP condition, subjects 3, 6, and 8 were also the ones obtaining the lowest overall scores, never reaching more than 80% of correct responses at any frequency. Discrimination of HP generally appeared to be best in a narrow mid-frequency range, which was different for each subject but always between 200 Hz and 800 Hz. When comparing the results to those from the HP detection experiment (section 2.3.1), one can notice that subjects 3 and 8 had difficulties in HP discrimination although they showed a very good detectability of HP, which suggests that some subjects with difficulties in frequency discrimination may detect binaural pitch very well but still find it difficult to hear the musical aspects of the pitch.

2.3.3 Melody recognition

The third experiment, similar to that carried out by Akeroyd *et al.* (2001), aimed at measuring the melody recognition capabilities of the subjects using pure tones as well as the six different binaural pitch types.

Method

The experiment consisted of 70 trials testing 7 different types of stimulus: pure tone, HP, BEP, BICEP, RBICEP, BBP, and BIBP (see Figure 2.2). For the HP, BBP, and BIBP configurations, the bandwidth of the phase shifted area was equal to 16% of the boundary frequency. For the BEP configuration, no transition range was used. 10 different melodies, each consisting of 16 equal-duration notes (Table 2.1(a)), were played using each type of stimulus. The frequency of the pure tone or the boundary frequency of the binaural pitch signal was equal to the frequency of each note (Table 2.1(b)). The order of the trials was randomized in such a way that neither the same stimulus type, nor the same melody was played in two successive trials. On each trial, the melody consisted of 16 notes of 300-ms duration, each separated by a silence of 100 ms. A 30-ms cosine-ramp was applied to the onset and offset of each note. In order to have equal-duration notes, the rhythm of the melodies required to be modified, so that long-duration notes were split into several equal-duration notes. A 10-AFC procedure was used. After the stimulus was presented once in each trial, the task of the subject was to indicate through a computer interface which melody was played. No feedback was provided. The experiment was repeated twice, *i.e.*, each subject participated in a total of 3 runs.

Table 2.1: The 10 melodies used in the melody recognition experiment, with corresponding notes and their frequencies.

(a) The 16 notes forming each of the 10 melodies.

		Notes															
	Melody	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Au Clair de la Lune	C5	C5	C5	D5	E5	E5	D5	D5	C5	E5	D5	D5	C5	C5	C5	C5
2	Frère Jacques	C5	D5	E5	C5	C5	D5	E5	C5	E5	F5	G5	G5	E5	F5	G5	G5
3	The Alphabet Song	C5	C5	G5	G5	A5	A5	G5	G5	F5	F5	E5	E5	D5	D5	C5	C5
4	Old MacDonald Had A Farm	F5	F5	F5	C5	D5	D5	C5	C5	A5	A5	G5	G5	F5	F5	F5	F5
5	London Bridge	G5	A5	G5	F5	E5	F5	G5	G5	D5	E5	F5	F5	E5	F5	G5	G5
6	Yankee Doodle	F5	F5	G5	A5	F5	A5	G5	C5	F5	F5	G5	A5	F5	F5	E5	E5
7	Sur le pont d'Avignon	F5	F5	F5	F5	G5	G5	G5	G5	A5	A5$^{\#}$	C6	F5	E5	F5	G5	C5
8	Chimes of Big Ben	A5	F5	G5	C5	C5	G5	A5	F5	A5	F5	G5	C5	C5	G5	A5	F5
9	Eurovision Theme	C5	F5	F5	F5	G5	A5	A5	F5	F5	C6	C6	C6	B5	A5	A5	A5
10	Jingle Bells	E5	E5	E5	E5	E5	E5	E5	E5	E5	G5	C5	D5	E5	E5	E5	E5

(b) Note frequencies.

Note	**Frequency**
C5	523.25 Hz
D5	587.32 Hz
E5	659.26 Hz
F5	698.46 Hz
G5	783.99 Hz
A5	880.00 Hz
A5$^{\#}$	932.33 Hz
B5	987.77 Hz
C6	1046.50 Hz

Table 2.2: Results from a set of Wilcoxon matched-pairs signed-ranks tests for each pairwise comparison of the pitch types used in the melody recognition experiment. Comparisons for the 13 normal-hearing subjects are given above the diagonal. Comparisons for the 8 hearing-impaired subjects are given below the diagonal. For a given pair, 'Yes' indicates that the results differ significantly, 'No' that no significant difference was found. Numbers in the brackets indicate the corresponding T and N values (T/N). A significance level of 0.05 was used.

	Tone	**HP**	**BEP**	**BICEP**	**RBICEP**	**BBP**	**BIBP**
Tone	—	No (18/9)	No (19/12)	Yes (0/11)	Yes (0/13)	No (5/7)	No (17/9)
HP	No (10.5/8)	—	No (4.5/8)	Yes (2/11)	Yes (0/13)	No (24/10)	No (17/9)
BEP	Yes (0/8)	Yes (1.5/8)	—	Yes (4/12)	Yes (0/13)	No (15/11)	No (10.5/8)
BICEP	Yes (0/8)	Yes (0/8)	No (6/8)	—	Yes (7.5/13)	Yes (1.5/11)	Yes (6/13)
RBICEP	Yes (0/8)	Yes (0/8)	No (3/7)	No (14.5/8)	—	Yes (0/13)	Yes (0/13)
BBP	No (4/7)	No (6/7)	Yes (0/7)	Yes (0/8)	Yes (0/8)	—	No (24.5/10)
BIBP	Yes (0/8)	Yes (1.5/8)	Yes (1/8)	Yes (0/8)	Yes (0/8)	No (10/8)	—

It was made sure before the experiment that the subjects had a sufficient knowledge of the melodies in order to avoid strong bias effects because they knew some of the melodies better than others. Some of the chosen melodies (no. 4, 5, 7, 9, and 10 in Table 2.1(a)) were different from those used by Akeroyd *et al.* (2001) so that they were more likely to be known by the listeners. Before they participated in the experiment, the subjects went through a learning procedure of the 10 chosen melodies. The learning process contained two phases: (1) The aim of phase 1 was to familiarize the subjects with the melodies themselves. The subjects could listen to the pure-tone version of the melodies as many times as they wished using a computer interface, through which they could enter their own title or comments in order to help them identify each melody. (2) The aim of phase 2 was to familiarize the subjects with hearing the melodies played with binaural pitch signals. Still using a computer interface, the subjects went through a fixed procedure where they were instructed to listen to the melodies a limited number of times. Melodies 1 to 10 were all played in 5 different conditions with pure-tone, HP, BEP, and BICEP stimuli. At the end of phase 2, the subjects had the opportunity to listen to the pure-tone melodies again, as in phase 1.

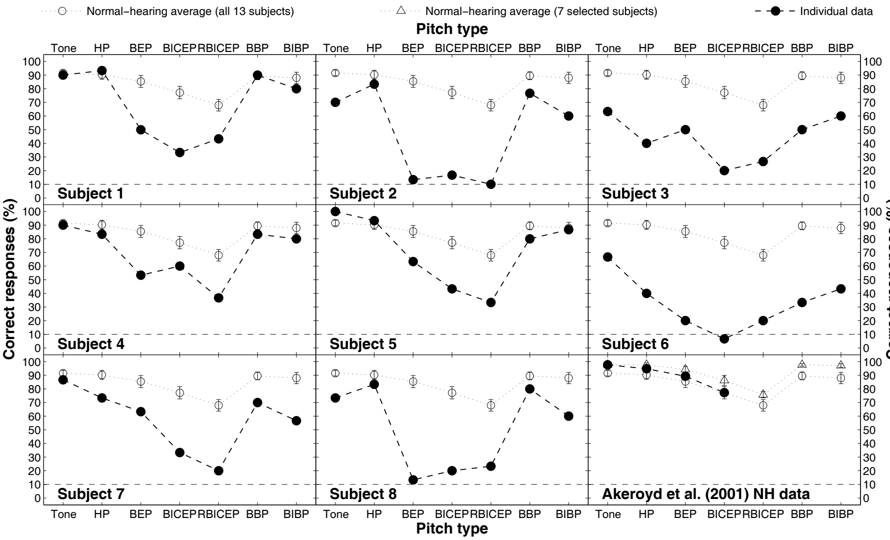


Figure 2.5: Melody recognition of pure-tones and six types of binaural pitches: average results and standard error over 13 normal-hearing listeners and individual results from 8 hearing-impaired listeners. The mean value of the three runs for each subject was used in the calculations. The ninth subplot in the lower-right corner indicates the results obtained by Akeroyd et al. (2001) with normal-hearing listeners.

Results and discussion

The melody recognition ability of the subjects was measured as the percentage of correct responses for each stimulus type. Chance level corresponded to 10% of correct responses. The average results and standard error over 13 NH subjects as well as individual results from all 8 HI subjects are shown in Figure 2.5. The results from 7 of the NH test subjects, selected for their very good knowledge of the pure-tone melodies after the learning phase, have also been averaged. The results shown in Figure 2.5 represent the average of the three runs performed by the subjects. In order to evaluate the statistical significance of the differences in the results for the seven stimulus types, a set of Wilcoxon matched-pairs signed-ranks tests (Fisher and van Belle, 1993) was performed. Results from the statistical tests are summarized in Table 2.2.

In their study, Akeroyd *et al.* (2001) found that, in NH listeners, the three binaural-pitch types they tested (HP, BEP, and BICEP) were less salient than pure tone signals, and that HP was the most salient binaural-pitch configuration, followed by BEP and BICEP. The average results over the four runs performed in their experiment are indicated in the lower-right subplot in Figure 2.5. Average results from the current study seem to follow the same trend as theirs, although melody recognition using HP and BEP does here not significantly differ from that obtained with pure tones (Table 2.2). HP actually seems to produce the same clear pitch perception as the pure tones, which is consistent with the results of the HP discrimination experiment (section 2.3.2), which suggested that HP was as musical as pure tones in the frequency range considered here. BICEP was found to be significantly less salient than BEP, which itself is on average less salient than HP. Results obtained with the three additional pitch types showed a clear pattern. RBICEP was found to be significantly less salient than all other pitch types. The subjects generally found the melodies more difficult to perceive with RBICEP than with BICEP. Hartmann and McMillon (2001) found that RBICEP was easier to perceive than BICEP for low boundary frequencies (below about 600 Hz), and that it was less salient than BICEP for high boundary frequencies (above about 600 Hz). Frequencies of the notes used in the 10 melodies were between 523.25 Hz and 1046.50 Hz, which means that most notes, and therefore most values of f_b, lie in the ‘high-frequency’ binaural pitch region. Therefore, in this test, RBICEP should then be slightly weaker than BICEP, which would explain the difference between both configurations in the results. The recognition obtained with the BBP and BIBP configurations appeared to be as good as with the pure-tone and HP configurations, although the phase-difference patterns in HP (linear phase shift), BBP (constant phase shift), and BIBP (random phase shift), are very different. This suggests that the pattern of the phase shift along frequency used in the area around f_b may not be of much importance for the auditory system when detecting binaural pitch, as long as the left and right noises are in phase outside this area. The *number of edges* (one or two) in the interaural phase-difference pattern appears to be what mostly determines the salience of the produced binaural pitch, with either simple phase-shift edges (sharp or smooth) or coherence edges. HP (2 smooth edges), BBP (2 sharp edges), and BIBP (2 coherence edges), all created by a pattern with two edges, are significantly more salient than BEP (1 simple edge), BICEP (1 coherence edge), and RBICEP (1 coherence edge), which are created by a pattern with only one edge. Whether the edges are *coherence* edges or not appears to be the secondary factor influencing the salience of the pitch.

For none of the HI subjects were the results similar to the NH results for all pitch types. Thus, all types of hearing impairment encountered in this study affected melody recognition produced by binaural pitch in some way. In order to make the results easier to read, the perception of each binaural-pitch type can at first be evaluated in a simple way by dividing the vertical axis of Figure 2.5 in three intervals:

1. *70%-100% of correct responses (good):* the pitch is easily detected by the subject and is musical, *i.e.*, can be used to produce recognizable melodies;
2. *30%-70% of correct responses (intermediate):* the pitch is detected by the subject but with a reduced salience or musicality;
3. *0%-30% of correct responses (poor):* the pitch cannot be detected by the subject or has a very weak salience or musicality.

With the HP configuration, all subjects obtained a good recognition, except subjects 3 and 6 whose scores were intermediate. The finding that none of the subjects had poor recognition is consistent with the finding that all of them were able to detect HP immediately in the preliminary test and obtained a good detectability of HP at least in the mid-frequency range in the HP detection experiment (see section 2.3.1). With BEP, no HI subject obtained a good recognition, and BEP was significantly less salient than HP for the HI group (Table 2.2). This shows a very important difference between HI and NH subjects, for whom BEP almost had the same strong salience as HP. Whereas 5 subjects were able to detect BEP, the other 3 subjects (2, 6, and 8) obtained scores that were very close to chance level. With BICEP and RBICEP, subjects 1, 4, and 5 obtained intermediate recognition for both pitch types, subject 7 obtained an intermediate recognition for BICEP and a poor recognition for RBICEP, and all other subjects had a poor recognition for both pitch types. Among the three subjects having intermediate scores, three of them (subjects 4, 5 and 7) found BICEP easier to perceive than RBICEP, which shows the same trend as observed with the NH subjects.

The scores obtained with BICEP and RBICEP were significantly lower than with HP for all subjects, and lower than with BEP for all subjects except subject 4. BICEP and RBICEP clearly appeared to be the two most difficult pitch types to perceive for the HI subjects, all scores lying below 50% except for subject 4. The three subjects who supposedly could not perceive BEP all had a poor recognition, suggesting that they also have a very weak or no perception of the binaural coherence edge pitches. Results obtained with BBP are very similar to those obtained with HP, as was also found for the NH subjects. However, while the results for HP and BIBP were similar for the NH listeners, BIBP appeared to be significantly less salient than HP for the HI group. The results were not correlated to the average value of the subjects' pure-tone hearing thresholds between 125 Hz and 2000 Hz, with a correlation coefficient[2] $r = -0.09$. A patient with a normal audiogram may indeed still suffer from some deficiency of the auditory system: subjects 1 and 2 had normal hearing thresholds and could then easily have been considered as being NH subjects, but the results they obtained in this experiment show that there must be deficits in their binaural processing of sounds.

The main results for the HI listeners can be summarized by dividing the subjects into three groups:

1. Listeners who can perceive *all types of pitches clearly*, but for whom binaural pitch generally is less salient than for NH persons. For these listeners, as in NH listeners, the most salient pitches are HP, BBP, and BIBP, followed by BEP, BICEP, and RBICEP. Subjects 1, 4, and 5 belong to this group.
2. Listeners who can perceive binaural pitches clearly *except when coherence edges are used* to generate them. These listeners perceive HP, BBP, and BEP as subjects from group 1, but find it very difficult to hear BICEP and RBICEP. Subjects 3 and 7 belong to this group.
3. Listeners who can perceive binaural pitches as those from group 2, *except with the BEP configuration*, which they find almost impossible to hear. Subjects

---

[2] The average of the results with all six binaural pitch stimuli in the melody recognition test and the average of the ERB values from both ears in the frequency selectivity test were used to derive the correlation coefficients. Correlation significance was determined with a significance level of 0.05.

2, 6, and 8 belong to this group. These three subjects all mentioned that they could not hear the melodies at all when they were played with BEP and BICEP stimuli.

2.4 Relation with measures of spectral and temporal resolution

The three following tests aimed at characterizing more precisely the hearing impairment of each HI subject, as well as investigating the correlation between the performance in the binaural pitch experiments and that in basic tasks of frequency discrimination, frequency selectivity, and temporal gap detection.

2.4.1 Frequency discrimination

This first additional test aimed at measuring the subjects' just noticeable difference (JND) in frequency of pure tones, HP, and BIBP, at a frequency of 500 Hz. All 10 HI listeners as well as 8 NH listeners participated in this experiment.

Method

The frequency JND was measured using a 3-interval, adaptive 3-AFC procedure. On each trial, two intervals had an identical frequency of 500 Hz, whereas the other randomly chosen interval had a frequency equal to $500(1+x/100)$ Hz, where x is the tracking variable expressed as a percentage of the test-frequency. The stimuli had a duration of 500 ms and were separated by 500 ms of silence. The task of the listener was to indicate through a computer interface in which interval the pitch was higher than in the two other intervals. Feedback was provided to the subject after each response. A 1-up 2-down tracking rule converging at the 70.7% point of the psychometric function (Levitt, 1971) was used. The starting value of x was 25%, and the step-size was decreased after each upper reversal, with successive reduction factors

Table 2.3: (a) Frequency just noticeable difference (JND) of pure-tone, HP, and BIBP stimuli expressed as a percentage of the (boundary) frequency of 500 Hz: average over 8 normal-hearing subjects and individual results from 10 hearing-impaired subjects. (b) Equivalent rectangular bandwidth (ERB), in Hz, of the auditory filter centered at 500 Hz for 6 normal-hearing and 9 hearing-impaired subjects (both ears). The ERB value suggested by Glasberg and Moore (1990) is 78.7 Hz. (c) Broadband noise gap detection threshold values (means and standard deviations), in ms, measured for 9 hearing-impaired subjects and 5 normal-hearing subjects at a presentation level of 75 dB SPL. The normal-hearing gap detection threshold as measured by Zeng *et al.* (2005) at a presentation level of 50 dB SPL is 3 ms.

	(a) Freq. JND at 500 Hz (%)			**(b) ERB at 500 Hz (Hz)**		**(c) Gap detection**
NH subjects	**Tone**	**HP**	**BIBP**	**Right ear**	**Left ear**	**Threshold (ms)**
Average	0.6	2.3	4.3	97.6	97.6	4.6 ± 0.6
Subject A	—	—	—	92.8	87.8	4.8 ± 0.6
Subject B	—	—	—	98.8	95.8	—
Subject C	—	—	—	110.8	122.6	—
Subject D	—	—	—	116.0	112.2	4.5 ± 0.7
Subject E	—	—	—	78.1	76.4	4.9 ± 0.6
Subject F	—	—	—	85.3	94.2	4.8 ± 0.5
Subject G	—	—	—	—	—	3.8 ± 0.6
HI subjects	**Tone**	**HP**	**BIBP**	**Right ear**	**Left ear**	**Threshold (ms)**
Subject 1	2.5	12.4	22.2	—	—	—
Subject 2	0.8	3.2	15.5	102.3	105.8	4.6 ± 1.1
Subject 3	23.1	28.2	30.9	104.5	115.5	4.4 ± 0.7
Subject 4	1.2	5.4	15.1	138.8	116.8	6.7 ± 0.9
Subject 5	0.6	2.2	3.0	100.5	105.0	6.2 ± 0.9
Subject 6	9.1	15.4	11.5	109.3	104.6	6.8 ± 0.9
Subject 7	0.7	6.4	25.9	129.4	123.0	3.9 ± 0.6
Subject 8	8.4	12.3	11.0	171.5	187.6	6.2 ± 0.6
Subject 9	29.9	—	—	157.8	>200.0	12.9 ± 2.1
Subject 10	17.4	—	—	187.1	188.1	10.7 ± 1.1

of 2, 1.5, and 1.125. One run stopped after 12 reversals, and the JND was calculated as the average peak value of x over the last 8 reversals. A total of 6 runs were performed, 2 with each type of stimulus.

Results and discussion

The frequency JND was measured for each stimulus type as a percentage of the test frequency. The average results over the 8 NH subjects as well as individual results from the HI subjects are given in Table 2.3(a).

The frequency JND of the 500 Hz pure tone was on average found to be below 1% (5 Hz) for all NH subjects. The average result over all subjects was 0.6%, corresponding to 2.9 Hz. This is slightly higher than the value of 1-2 Hz generally obtained at 500 Hz in most experiments measuring the frequency JND (Moore, 2003). With the HP stimulus, the JND was up to 2.3% (11.7 Hz) on average, and reached 4.3% (21.4 Hz) on average with the BIBP stimulus. The JND value obtained for HP is relatively large compared to that found by Hartmann (1993) in his pitch matching experiment, probably due to the rather large width of the HP boundary region used here (16% of f_b). These results suggest that the pitch frequency of binaural pitches is not as precise as that of pure tones, which is consistent with the perception reported by some listeners that binaural pitches sound more like narrow-band noise signals than tones (especially BEP, BICEP, and RBICEP stimuli). The presence of coherence edges in the BIBP stimulus does spread even more the pitch frequency around f_b. However, in all three cases, the obtained frequency JNDs were small enough, *i.e.*, less than 9% or 1.5 musical semitones as used in the HP discrimination experiment (see section 2.3.2), to suggest that all three pitches are musical around 500 Hz.

For the HI subjects, 4 subjects showed results similar to the NH average (subjects 2, 4, 5, and 7) in the pure-tone condition. The other 6 subjects obtained higher JNDs than NH listeners, ranging from 2.5% for subject 1 to 8-10% for subjects 6 and 8, and up to 23% for subject 3. Note that the two subjects who could not perceive binaural pitch (9 and 10) obtained considerably higher JNDs for pure tones than the NH average. In the HP condition, only subjects 2 and 5 showed similar results to NH listeners, and all other subjects obtained higher JND values, ranging from 5% for subject 4 to 28% for subject 3. In the BIBP condition, only subject 5 showed a very good frequency discrimination, even compared to the NH average. All other subjects obtained much higher JND values, ranging from 11% for subject 8 to 31% for subject 3. For subjects 1, 2, 3, 4, 5, and 7, frequency discrimination was easiest for pure-tone stimuli, and their JNDs were higher in the HP condition than in the pure-tone condition, and higher in the BIBP condition than in the HP condition. For subjects 6 and 8, the HP stimuli were the most difficult to discriminate, followed by BIBP and pure-tone stimuli.

The frequency discrimination abilities of the subjects were not related to audibility as reflected in their audiogram. For example, subject 3 had an audiogram showing a very mild impairment, whereas its abilities to discriminate between frequencies appeared to be severely affected. The opposite is true for subject 5, who has very high hearing thresholds at high frequencies (above 80 dB HL), but is actually the only subject who showed a normal frequency discrimination for the three stimuli. Moreover, these two subjects have a noise-induced hearing loss, which also suggests that two hearing impairments from the same origin can lead to very different perceptual problems. The two subjects who had a normal audiogram (1 and 2) obtained different results for frequency discrimination. Subject 2 obtained comparable results to NH listeners with the pure-tone and HP stimuli, but had more difficulties in the BIBP condition. Subject 1 had a slightly higher JND than NH listeners in the pure-tone condition, but had obvious difficulties in discriminating both HP and BIBP.

The results are consistent with those obtained in the HP discrimination and melody recognition tests (sections 2.3.2 and 2.3.3), in which subjects 3, 6, and 8 showed the largest difficulty in discriminating pure-tone and HP stimuli, as well as recognizing pure-tone melodies. These three subjects are also the ones with the highest frequency JNDs for pure tones. This confirms our expectation that problems with frequency discrimination would potentially affect the musical abilities of some subjects.

2.4.2 Frequency selectivity

The second additional test aimed at measuring the frequency selectivity abilities of the subjects. The bandwidth of the auditory filter centered at 500 Hz, the frequency at which HP was found to be the most salient for all listeners (Figure 2.3), was evaluated monaurally in both ears of 9 HI listeners as well as 6 NH listeners.

Method

The notched-noise method (Patterson, 1976) was used for determining auditory filter shapes. The signal was a 500 Hz tone and the spectrum level of the noise was 40 dB. The detection threshold for the signal tone was measured at six relative notch widths $g = \Delta f / f_c$: 0, 0.05, 0.1, 0.2, 0.3, and 0.4, where Δf is the notch width and f_c the signal frequency. The lower and upper cut-off frequencies of the noise were equal to $f_c(0.6 - g)$ and $f_c(1.4 + g)$, respectively. A 3-interval, adaptive 3-AFC procedure was used for each threshold measurement. In each trial, the three intervals contained the notched noise, one randomly chosen interval also contained the tone signal. The subject's task was to detect the interval containing the pure tone. The initial level of the pure tone was 70 dB SPL. A 1-up, 2-down tracking rule with 4 step sizes of 8 dB, 4 dB, 2 dB, and 1 dB was used. The step size was halved after each second reversal. Using the average value over the last 6 reversals, the mean threshold value was calculated as a function of the relative notch width g for each subject. A rounded-exponential filter (Patterson *et al.*, 1982) was fitted to the experimental data using a least-square fit. The equivalent rectangular bandwidth (ERB) of the subject's auditory filter at 500 Hz was then derived from the parameters of the fitted filter.

Results and discussion

The ERB values obtained for both ears of each subject are given in Table 2.3(b). Although the 'normal' ERB value at 500 Hz, as suggested by Glasberg and Moore (1990), is equal to 78.7 Hz, the results obtained with the NH listeners of the present study showed a high variability among subjects, with ERB values ranging from 76 to 122 Hz. It is therefore not possible to consider the HI subjects who obtained ERB values in that interval in both ears (subjects 2 to 6) as having impaired frequency selectivity compared to the NH listeners of the present study. However, all other HI subjects did show a reduced frequency selectivity. The right-ear auditory filters of HI subjects 4 and 7 (ERB = 139 and 129 Hz) appeared to be slightly broader than normal, while HI subjects 8 to 10 obtained ERB values that were considerably higher than those of all other subjects. The two subjects who could not perceive any binaural pitch at all (HI subjects 9 and 10) are among those with the highest ERB values. However, the correlation between the ERB values obtained for HI subjects and their performance in the melody recognition experiment is not significant, with a correlation coefficient[2] $r = -0.63$.

The differences in interaural correlation as a function of frequency, inherent to binaural pitch stimuli, are essential to create binaural pitch. Because the bandwidth of the auditory filters has a direct influence on the internal representation of the interaural-correlation spectrum, an impaired frequency selectivity should limit the

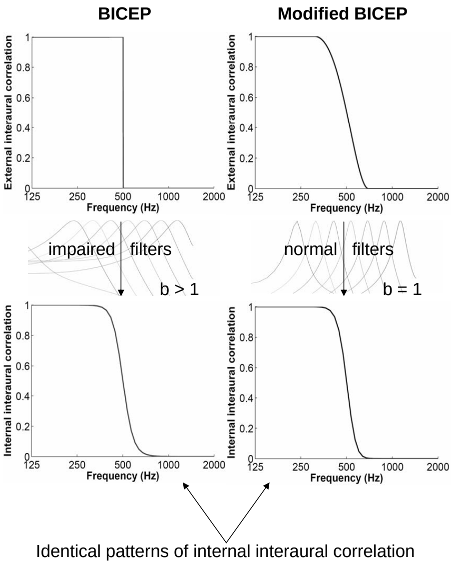


Figure 2.6: Modification of the BICEP stimulus using interaural correlation in order to simulate its perception by a listener with broader auditory filters.

ability of the listener to perceive binaural pitch. For instance, for the BICEP stimulus, the waveform interaural correlation (before auditory filtering) is equal to 1 below f_b and 0 above f_b (Figure 2.6). Once passed through a set of auditory filters, this external waveform interaural-correlation spectrum is transformed into an internal interaural-correlation spectrum in which the decay around f_b is not sharp anymore but progressive (Figure 2.6). The edge in this internal interaural-correlation spectrum becomes smoother if the auditory filters are broadened. If one assumes that the internal interaural-correlation representation is a key element used by the binaural system in the processing of binaural signals, it might then be possible to further investigate the influence of auditory-filter broadening on binaural pitch perception by modifying the external waveform interaural-correlation spectrum of the stimulus itself. The external interaural-correlation spectrum of the BICEP stimulus was modified here such that the internal interaural-correlation spectrum it produced when fed through normal auditory filters matched the internal interaural-correlation spectrum produced by the original

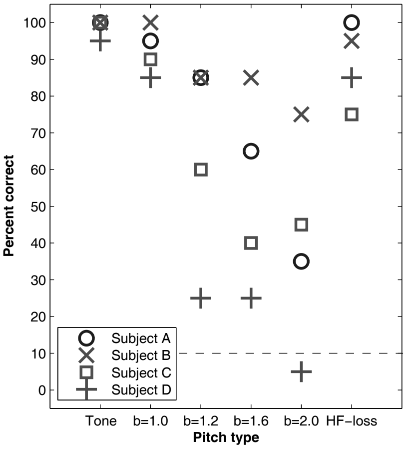


Figure 2.7: Melody-recognition of pure-tones, BICEP and 4 modified versions of the BICEP stimulus in 4 normal-hearing listeners. b = 1.0 : BICEP; b = 1.2 - 1.6 - 2.0 : modified BICEP (simulating broader auditory filters); HF-loss: BICEP with reduced amplitude at high frequencies (simulating a higher threshold). The equivalent rectangular bandwidths (ERB) of the left and right auditory filters centered at 500 Hz for the same 4 subjects are indicated in Table 2.3(b).

BICEP stimulus when fed through a set of impaired auditory filters (Figure 2.6). This was done by introducing a transition in the external interaural-correlation spectrum instead of a sharp edge. In this way, it was possible to simulate the effect of broadened auditory filters and to measure the consequences of this on binaural pitch perception in NH listeners. The gammatone filterbank (Patterson *et al.*, 1995) provided in the auditory toolbox by Slaney (1998) was used to simulate auditory filtering, and the broadening factor *b* was varied to simulate different filter bandwidths.

The melody recognition ability of 4 NH subjects was measured with pure tones, BICEP stimuli, and modified versions of the BICEP with broadening factors of 1.2, 1.6, and 2.0. The procedure was identical to that described in section 2.3.3. The experimental results are given in Figure 2.7. The melody recognition abilities of the subjects were found to decrease progressively with increasing values of *b*, which suggests that the bandwidth of the auditory filters might play an important role in the ability of a listener to perceive BICEP in a musical way. The strength of this effect was different for the different subjects. Interestingly, a similar variability was also found when estimating the frequency selectivity of the same 4 subjects at 500 Hz (Table 2.3(b)). The two subjects whose melody recognition abilities decreased the most rapidly with increasing b (subjects C and D) were also those with the highest ERB values at 500 Hz, which supports the idea that performance with frequency selectivity and melody recognition of BICEP might be related.

In an additional stimulus configuration, reduced audibility at high frequencies was also simulated by attenuating the amplitude of the high-frequency (HF) components of the original BICEP stimulus. The amplitude of frequency components above 500 Hz was attenuated progressively with increasing frequency, with reductions of 5 dB at 1 kHz, 35 dB at 2 kHz, 65 dB at 3 kHz, and 80 dB at 4 kHz. This modification did not appear to have any effect (‘HF-loss’ in Figure 2.7), *i.e.*, the results obtained with the ‘HF-loss’ configuration were similar to those obtained with the original BICEP.

2.4.3 Temporal resolution

The third additional test aimed at measuring the temporal resolution of 9 HI and 5 NH listeners, using a gap detection threshold measurement.

Method

Temporal resolution was measured by estimating the detectability of temporal gaps in broadband noise. A fixed level of 75 dB SPL was used in this experiment. A 3-interval, adaptive 3-AFC procedure was used. All intervals contained white noise with a bandwidth of 3000 Hz and an upper cut-off frequency of 4000 Hz. One random interval contained a gap in the noise. A 1-up, 2-down tracking rule was used. The initial gap duration was 120 ms, and three step sizes corresponding to a reduction factor of 2, 1.5, and 1.125 were used. The step size was reduced after each second reversal. The gap detection threshold was calculated as the average over the last 8 reversals. A total of three runs was performed by each subject.

Results and discussion

The gap detection thresholds obtained for each subject are given in Table 2.3(c). Plomp (1964) and Penner (1977) proposed a typical threshold value of 2-3 ms for NH listeners, while Zeng *et al.* (2005) found the NH threshold to be about 3 ms. The values obtained in the present study were slightly higher, with an average value of 4.6 ms for NH listeners. HI subjects 2, 3, and 7 obtained results close to the NH average, while HI subjects 4, 5, 6, and 8 obtained slightly higher thresholds between 6 and 7 ms. The two HI subjects who could not perceive binaural pitch at all (HI subjects 9 and 10) obtained considerably higher thresholds than all other subjects. Overall, there is a significant correlation between the gap detection thresholds obtained for HI subjects and their performance in the binaural pitch experiments, as indicated by a correlation coefficient[2] $r = -0.79$. Furthermore, comparing the gap detection results with the pure-tone hearing threshold of the subjects, averaged over all audiometric test frequencies, gives a significant correlation coefficient[2] $r = 0.74$. This seems consistent with earlier studies on gap detection. HI subjects often show reduced temporal resolution as a result of the low sensation level of the stimuli and the reduced audible bandwidth of the stimuli (*e.g.*, Florentine and Buus, 1984; Salvi and Arehole, 1985).

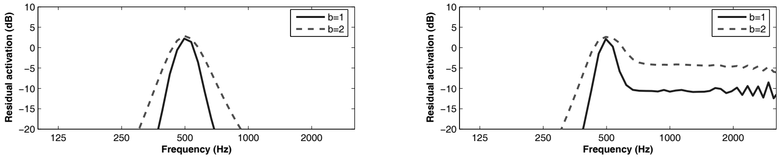
As indicated earlier, reduced frequency selectivity could affect binaural pitch perception already at a cochlear stage. It is therefore difficult to separate the effects caused by deficiencies in temporal processing from those caused by impaired frequency selectivity if the psychophysical experiments are performed in subjects who also have broadened auditory filters. It would be interesting to investigate binaural pitch perception in a group of subjects with similar frequency selectivity (*e.g.*, subjects with similar outer hair-cell functionality) but with different aspects of temporal deficiencies at cochlear or more central stages of the auditory system.

2.5 General discussion

2.5.1 Summary of the main results

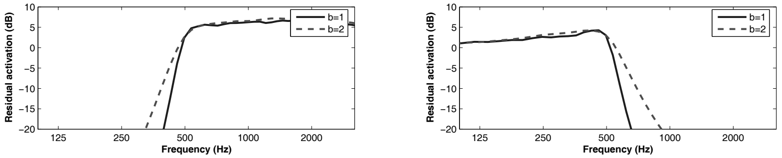
First, results for the normal-hearing listeners are summarized. A preliminary test with the Huggins' pitch (HP) showed that it produced an immediate sensation. In the first experiment, the detectability of HP was measured as a function of frequency, and the pitch was found to be most salient for boundary frequencies between 200 Hz and 1250 Hz, with a declining detectability below 200 Hz and above 1250 Hz. The second experiment was a pitch discrimination test and aimed at measuring the musicality of HP. The pitch was found to be as musical as that produced by pure tones for boundary frequencies between 200 Hz and at least 800 Hz. Discrimination of HP appeared to be progressively more difficult below 200 Hz and above 800 Hz. In the third experiment, the melody recognition abilities of the listeners were measured with pure tones and six different types of binaural pitches. It was found that the salience of binaural pitch was largest for HP, followed by the binaural edge pitch (BEP) and the binaural coherence edge pitch (BICEP). Using a constant or random phase-shift around the boundary frequency instead of the linear phase-shift of HP did not affect the salience and musicality of the produced pitch. In the case of BICEP, the pitch strength was found to be higher if coherent noise was used above the edge than below the edge, for edge frequencies between 520 Hz and 1050 Hz. The just noticeable difference in frequency (JND) at 500 Hz was measured for pure-tone, HP, and BIBP stimuli. The average JND was found to be higher for HP than for pure tones, and higher for the binaural incoherent-band pitch (BIBP) than for HP. The JNDs were found to be low enough to allow the identification of melodies played with those three stimuli for pitch frequencies around 500 Hz. Estimates of auditory filter bandwidth at the same frequency using the notched-noise method showed a surprisingly large variability among the normal-hearing listeners. A test simulating the effect of broadening of the auditory filters by modifying the interaural-correlation spectrum showed that binaural pitch perception became more difficult as the simulated bandwidth of the filters was increased. The strength of the effect was not the same among subjects, and subjects who already had relatively broad auditory filters performed worse if the smoothed interaural-correlation spectrum was introduced. The simulation of reduced audibility did not affect binaural pitch perception.

The same set of experiments was also carried out by hearing-impaired (HI) listeners with different types of auditory deficiencies. It was found that the HI listeners who reported central auditory processing deficits could not perceive binaural pitch at all. However, for most HI listeners, binaural pitch was perceivable, and in all cases produced an immediate sensation as in the normal-hearing (NH) listeners. Hearing impairment did not affect the detectability of HP, and the range of boundary frequencies in which the pitch is easily detectable was the same for some HI listeners as for NH listeners. However, for several HI listeners the frequency region in which the pitch was most salient was reduced on the low-frequency side or the high-frequency side, or on both sides. Discrimination of HP stimuli with close boundary frequencies was found to be typically more difficult for HI listeners than for NH listeners. In contrast to NH listeners for whom all binaural-pitch types were shown to be musical, the melody recognition abilities of HI listeners appeared to be very different depending on the pitch type used. Despite the fact that they could detect HP easily, some HI listeners only showed a very weak or non-existing perception of BICEP. Among these, some also showed a very weak or non-existing perception of BEP. The JND in frequency at 500 Hz was found to be considerably higher for some HI subjects than the NH average, and the same general trend was observed for both subject groups: For most subjects BIBP was more difficult to discriminate than HP, which was itself more difficult to discriminate than pure tones. The difficulties of HI subjects with frequency discrimination were consistent with their discrimination and melody recognition performances in the binaural pitch experiments. Results from the frequency selectivity test were not significantly correlated to those from the melody recognition experiment, but subjects who could not hear binaural pitch at all were also those who had the broadest auditory filters at 500 Hz. The largest gap detection thresholds were obtained by the subjects who could not perceive binaural pitch at all. Performance in the gap detection experiment was overall correlated to melody recognition with binaural pitch stimuli, but appeared to be influenced by the reduced sensation level or the reduced audible bandwidth of the stimuli by the subjects.



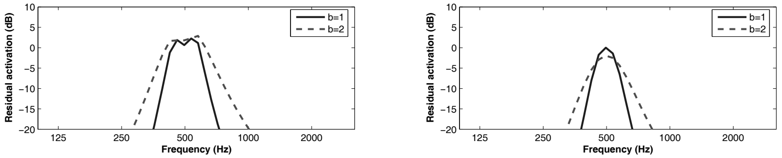
(a) Huggins' pitch (HP)

(b) Binaural edge pitch (BEP)



(c) Binaural coherence edge pitch (BICEP)

(d) Reversed BICEP (RBICEP)



(e) Binaural band pitch (BBP)

(f) Binaural incoherent-band pitch (BIBP)

Figure 2.8: Residual activation spectra recovered by the modified equalization-cancellation model for six types of binaural pitch signals with a boundary frequency of 500 Hz. b = 1: normal auditory filters; b = 2: broader auditory filters by a factor of 2. (a) Huggins' pitch. (b) Binaural edge pitch. (c) Binaural coherence edge pitch. (d) Reversed binaural coherence edge pitch. (e) Binaural band pitch. (f) Binaural incoherent-band pitch.

2.5.2 Modeling binaural pitch

The experimental results of the present study can be discussed in the light of the modified equalization-cancellation (mEC) model, as described by Culling *et al.* (1998). In this model, the left and right stimuli are fed through a bank of gammatone filters (Patterson *et al.*, 1995) and a model simulating hair-cell transduction (Meddis *et al.*, 1990). Each frequency channel is then treated independently during an EC process in which the RMS values of the left and right signals are first equalized and an interaural time delay τ giving an optimal cancellation is introduced. A spectrum of residual activation is obtained after the cancellation process. Using the auditory toolbox by Slaney (1998), the binaural-pitch signals used in the melody recognition experiment (Figure 2.2) were fed through the model for a boundary frequency of 500 Hz. Signals were 300 ms long and 80 frequency channels between 100 Hz and 22050 Hz were used. For the HP, BBP, and BIBP configurations, the bandwidth of the phase shifted area was equal to 16% of the boundary frequency. For the BEP configuration, no transition was used, as was indicated in Figure 2.2(b). The obtained residual-activation spectra for 5 averaged iterations of the model are plotted in Figure 2.8.

In the case of HP, BBP, and BIBP, where the left and right noises are in phase outside the frequency range close to f_b, the cancellation can be performed almost perfectly using a zero time delay ($\tau = 0$ ms), except in the area around f_b, where a complete cancellation cannot be reached for any value of τ. This gives rise in all three cases to a peak in the residual-activation spectrum (Figures 2.8(a), 2.8(e), and 2.8(f)). It is this peak, centered on the boundary frequency, which is assumed to account for the presence of the binaural pitch according to the mEC model. The model predicts almost no difference in perception between these three stimuli, because the residual-activation spectra are almost identical, a similarity which was also observed by Akeroyd and Summerfield (2000) for HP and BIBP. This is consistent with the perceptual results from the melody recognition experiment (section 2.3.3), where scores obtained by the listeners were very similar for HP, BBP, and BIBP. For the BEP signal (Figure 2.8(b)), cancellation can only occur almost completely below the boundary frequency, where the left and right signals are in phase, whereas cancellation is only partially achieved above f_b, where the signals are in antiphase. This is due to the fact that, for noise signals, a single interaural phase delay corresponds to slightly different interaural time delays at several frequencies close to each other. Within the mEC model that uses a single value of the time delay τ per frequency channel, this means that cancellation cannot be obtained for all frequencies within a single channel (Culling *et al.*, 1998). The presence of a peak is due to the fact that frequency channels close to f_b admit frequency components with a wide range of different time delays, thus making cancellation more difficult (Culling *et al.*, 1998). Here again, it is this peak which is assumed to account for the presence of a binaural pitch. The fact that cancellation is poorer on one side of f_b with the BEP than with the HP stimulus could explain why the salience of BEP is generally found to be weaker than that of HP. Because BBP is in fact similar to the combination of a BEP stimulus and a reversed BEP stimulus with close boundary frequencies, the peak obtained with the BBP stimulus looks like the combination of two neighboring peaks which would be produced by BEP stimuli, but are not resolved here due to the narrow bandwidth of the phase-shifted area of BBP. In the case of the BICEP and RBICEP configurations (Figures 2.8(c) and 2.8(d)), cancellation occurs almost perfectly on the coherent side of f_b, but becomes very difficult on the incoherent side, due to the random phase shifts introduced. The perception of a binaural pitch in these cases can only be explained by the presence of an edge in the residual-activation spectrum, which means that pitch detection could be produced by the enhancement of such an edge by the auditory system. The contrast enhancement at the edge could be due to a process of lateral inhibition similar to that possibly responsible for perception of a weak pitch near the edge frequency of a high-pass or low-pass noise (Klein and Hartmann, 1981; Culling *et al.*, 1998). Akeroyd *et al.* (2001) demonstrated that adding a simple contrast enhancement step to such a model could predict the differences in salience between HP, BEP, and BICEP. However, such a process is not consistent with the conclusions of Hartmann (1984), who found no evidence for the existence of central lateral inhibition.

The presence of edges in the interaural phase-difference pattern could be what mostly influences the salience of the produced binaural pitch. A parallel may be drawn between the importance of the number of edges in the phase-difference pattern and in the residual-activation pattern resolved by the mEC model. The most salient pitches are indeed those which give rise to a peak of residual activation (HP, BBP, BIBP), followed by the BEP, which creates a peak but with a major decay only on one side, and the least salient are those which only create an edge in the residual-activation spectrum (BICEP, RBICEP). Moreover, if an EC process is performed by the auditory system, the results obtained with HI subjects in the melody recognition experiment then suggest that subjects who could not perceive BICEP and RBICEP (groups 2 and 3, see section 2.3.3) had difficulties detecting an edge, but not a peak, in the residual-activation spectrum. If it is assumed that the auditory system performs a contrast enhancement process in order to detect the edges in the spectrum, it could then mean that these subjects could not perceive BICEP and RBICEP because of some deficiency in the mechanism underlying lateral inhibition.

It can be seen in Figure 2.8 that the width of the auditory filters affects the sharpness of the peak or edge, but hardly affects its amplitude (dashed curves in Figure 2.8). Peak amplitude does therefore *not* seem to account correctly for the salience of the binaural pitch in this model, which is not consistent with current hypotheses that pitch salience is determined by the amplitude of the residual activation (Culling *et al.*, 1998). Besides its amplitude, the sharpness of the peak and the amount of cancellation achievable by the system on both sides of the peak seem to be two parameters related to the salience of the binaural pitch. This favors a pitch detection process in which the difference in residual activation between neighboring frequency channels would play the most important role. A process of contrast enhancement, which involves such an across-frequency comparison, has been suggested in earlier studies (*e.g.* Akeroyd *et al.*, 2001). As an alternative approach to the purely temporal processing assumed by the mEC model, a model involving the direct spatial cross-correlation of non-corresponding left and right frequency channels might also be interesting to consider (Loeb, 1983; Shamma *et al.*, 1989; Shamma, 2001).

2.5.3 Implications and conclusions

The fact that the binaural pitch perception of hearing-impaired listeners was found not to be related to the shape of their audiograms confirms that additional tests are essential to better characterize hearing impairment and to determine which functions of the auditory system are affected. Frequency selectivity is a very interesting measure, and the width of the auditory filters was found to have an important influence on the perception of binaural stimuli, and to explain some differences observed even among normal-hearing listeners. The fact that different hearing-impaired listeners were found to react differently to binaural-pitch stimuli and that their perception was either immediate or non-existing makes binaural pitch stimuli interesting and useful also for clinical diagnostics. For example, if it was shown that only listeners with central auditory deficiencies were unable to perceive binaural pitch, a short binaural pitch detection test could be designed to immediately provide information about the presence of a deficiency in the upper stages of the auditory system. This would be consistent with the suggestion by Dougherty *et al.* (1998) that binaural pitch signals could possibly be used to detect cognitive disorders such as dyslexia. Further studies with groups of hearing-impaired subjects suffering from similar impairments would be relevant in order to evaluate if such clinical applications can be implemented. Experiments with homogeneous subject groups would also be essential to study the influence of a deficiency in a specific area of the auditory system on binaural pitch perception.

3

Relating binaural pitch perception to the individual listener's auditory profile[†]

The ability of eight normal-hearing listeners and fourteen listeners with sensorineural hearing loss to detect and identify pitch contours was measured for binaural-pitch stimuli and salience-matched monaurally-detectable pitches. In an effort to determine whether impaired binaural pitch perception was linked to a specific deficit, the auditory profiles of the individual listeners were characterized using measures of loudness perception, cognitive ability, binaural processing, temporal fine structure processing and frequency selectivity, in addition to common audiometric measures. Two of the listeners were found not to perceive binaural pitch at all, despite a clear detection of monaural pitch. While both binaural and monaural pitches were detectable by all other listeners, identification scores were significantly lower for binaural than for monaural pitch. A total absence of binaural pitch sensation coexisted with a loss of a binaural signal-detection advantage in noise, without implying reduced cognitive function. Auditory filter bandwidths did not correlate with the difference in pitch identification scores between binaural and monaural pitches. However, subjects with impaired binaural pitch perception showed deficits in temporal fine structure processing. Whether the observed deficits stemmed from peripheral or central mechanisms could not be resolved here, but the present findings may be useful for hearing loss characterization.

[†] This chapter is based on Santurette and Dau (2011b).

3.1 Introduction

Binaural pitch, also often termed dichotic pitch, is an auditory pitch sensation which can arise when two noise stimuli are presented simultaneously to the left and right ear, with an interaural phase shift over a specific frequency band (Cramer and Huggins, 1958). Such stimuli differ from most pitch-evoking stimuli in the sense that they require binaural interaction before any information relevant for pitch extraction can be available to the auditory system. The fact that broadband white noise, which has a flat power spectrum and does not evoke a pitch when presented monaurally, can be used to create a binaural pitch sensation, implies that such a pitch is formed centrally, using fine temporal disparities between the left and right peripheral channels. Therefore, a number of suggested models make use of interaural time differences to account for the existence of binaural pitch. These include equalization-cancellation models (*e.g.*, Durlach, 1960; Culling *et al.*, 1998), in which binaural cancellation occurs after the amplitude and phase of the left and right peripheral signals have been equalized, and central-spectrum models (*e.g.*, Raatgever and Bilsen, 1986), in which a central activity pattern is computed along the two dimensions of frequency and internal interaural time delay.

Despite its relatively small salience, binaural pitch was found to be immediately perceiveable and allow melody recognition in normal-hearing (NH) listeners (Akeroyd *et al.*, 2001). In hearing-impaired (HI) listeners, binaural-pitch perception was found to be either as immediate as in NH listeners or totally absent (Santurette and Dau, 2007). Interestingly, the two subjects who could not perceive binaural pitch at all in the latter study were the ones for whom deficits were likely to be present in central areas of the auditory system. Therefore, it was suggested that binaural pitch stimuli could be relevant to use in order to help characterizing hearing loss in individual patients, because such stimuli are easy to generate and to include in a fast pitch detection test. However, a clear link between impaired binaural pitch perception and a specific auditory deficit or hearing loss type needs to be established before a binaural pitch test may be considered useful.

In addition to reduced sensitivity, sensorineural hearing loss is often accompanied by the alteration of several important properties of the normal auditory system. First, most cases of cochlear hearing loss involve damage in outer hair cells (OHCs), whose active mechanism is crucial for a sharp frequency selectivity (Ruggero and Rich, 1991). Asymmetries in OHC damage may also lead to differences in auditory-filter outputs in the two ears. This could affect binaural pitch perception by reducing the interaural correlation in internal noise representations, thereby making binaural unmasking processes less efficient (Staffel *et al.*, 1990). Second, damage to inner hair cells (IHCs) and auditory-nerve fibers may affect the acuity of temporal fine structure (TFS) coding in HI listeners. Abnormalities in, *e.g.*, the propagation time of the basilar-membrane travelling wave and neural phase-locking mechanisms, may contribute to impaired TFS processing (Moore, 2007). Such deficits, as well as ear asymmetries in TFS processing, may lead to a degraded representation of interaural phase differences (IPDs), and thus affect binaural pitch perception. Additionally, deficits in more central mechanisms involved in the integration and processing of binaural information, at or above the level of the medial superior olive, may have deleterious effects on the ability to perceive binaural pitch. Ultimately, accurate binaural pitch perception should also rely on well-functioning central pitch extraction mechanisms.

In a first investigation of the effects of hearing loss on binaural pitch perception, Santurette and Dau (2007) showed that the detection and melody recognition scores of HI listeners with binaural pitch stimuli were not correlated with low-frequency pure-tone hearing thresholds. Their findings also demonstrated that impaired frequency selectivity affected pitch salience, but that estimates of auditory-filter bandwidths were not correlated with melody recognition abilities with binaural pitch stimuli. Nitschmann *et al.* (2010) also showed that, even though HI listeners who were unable to perceive binaural pitch had elevated auditory filter bandwidths, other HI listeners with similar bandwidths showed an immediate binaural pitch sensation. Therefore, impaired frequency selectivity alone cannot explain the inability of some listeners to perceive binaural pitch.

As no spectral cues are present in the physical stimulus, binaural pitch perception can be expected to rely heavily on accurate TFS processing and intact binaural integration mechanisms. However, so far it has not been investigated whether the inability of some HI listeners to perceive binaural pitch reflects a deficit in binaural processing *per se*, which would be located at the level of the brainstem or at higher auditory stages, or if reduced temporal acuity in the periphery is sufficient to account for impaired binaural pitch perception. The results of Strelcyk and Dau (2009), who showed a significant correlation between binaural and monaural measures of TFS processing in HI listeners, suggested that an impaired monaural representation of TFS might account for the binaural deficits. Moreover, TFS processing outcomes were not correlated to measures of frequency selectivity in their study. In addition, Nitschmann *et al.* (2010) found that the ratio of binaural to monaural estimates of frequency selectivity did not differ between NH and HI listeners, suggesting that the deficits apparent in binaural measures reflect a basic peripheral auditory impairment, rather than a specific binaural impairment.

The present study related the ability of HI listeners to process binaural pitch stimuli to their performance in an extensive set of specific measures of basic auditory and cognitive functions, referred to as the “auditory profile” in the following. A first aim was to clarify whether specific deficits are found in listeners with absent or reduced binaural pitch perception, and which basic functions are most crucial for a salient binaural pitch sensation to arise. A second aim was to investigate the relationship between the different auditory-profile measures as such, as an analysis of correlations between the outcomes of the different tests might help better understand the auditory processes involved in sensorineural hearing loss, and at which level of the auditory system they may take place. Altogether, the study of similarities and discrepancies in the performance of individual HI listeners in the different tests may also indicate which measures are informative and which are redundant when attempting to fully characterize hearing loss. Here, the approach was to perform an extensive set of tests on a heterogeneous group of HI listeners, with the postulate that pointing out individual differences might be as informative as focusing on group averages. While the tested subject group may reflect the diversity of impairments found in the HI population, this approach does not provide a direct assessment of the diagnostic value of a binaural pitch test in a clinical setting. Instead, it was attempted here to determine whether binaural pitch is an informative tool that may complement or replace other existing measures when attempting to characterize hearing loss.

Binaural pitch perception was evaluated in two experiments (section III). The first short test involved detection of a musical scale played with binaural pitch stimuli. The second more extensive test evaluated detection and identification of pitch contours played either with binaural pitch stimuli, or with pitch-evoking noise stimuli for which the pitch was detectable monaurally. In addition to being a more accurate measure of binaural-pitch detection, the comparison of binaural and monaural identification scores in this second test allowed to make the distinction between a binaural deficit and a more general difficulty in extracting tonal objects from noise (Chait *et al.*, 2007).

The test battery used to characterize the listeners' auditory profiles was designed to evaluate basic functions, which may be affected by hearing loss. The battery included the common audiometric measures (section IV): pure-tone air-conduction and bone-conduction thresholds, tympanograms and stapedius reflex responses to test for the presence of middle-ear dysfunction and potential auditory or facial nerve disorders, and click-evoked otoacoustic emissions to further assess cochlear function. Loudness perception estimates (section V) were also included, whereby a set of loudness curves was obtained using a categorical loudness scaling procedure. This was used to determine adequate presentation levels in other tests, but also in order to evaluate recruitment. The loudness growth measures were supplemented with a standard intensity-increment detection test, to better assess the presence of cochlear and retrocochlear disorders. The cognitive abilities of the listeners were evaluated (section VI) using a lexical decision task and a reading span test, in order to estimate the possible influence of non-auditory deficits on the different tasks. The ability of the listeners to take advantage of binaural processing when extracting signals from noise was then evaluated, using two measures of binaural unmasking (section VII): Binaural masked detection thresholds of sinusoidal tones in band-limited noise were compared for homophasic and antiphasic tones, and the spatial release from masking was evaluated in monaural and binaural conditions for speech reception thresholds (SRT) in speech-shaped noise. The latter measure was chosen because it did not only allow for evaluation of binaural processing, but also of speech intelligibility and spatial release from masking, within a single experiment. The ability of the listeners to use TFS cues in quiet was tested both binaurally and monaurally (section VIII): Binaural TFS processing abilities were evaluated *via* detection of an IPD in a sinusoidal tone as a function of frequency, and frequency-modulation (FM) detection at a low FM-rate was used as a measure of monaural TFS processing. Finally, monaural frequency selectivity was investigated and the left and right auditory filter shapes were estimated from the results of a notched-noise experiment (section IX). A detailed description of all tests is provided in the appendix of this chapter.

3.2 Methods

3.2.1 Subjects

8 normal-hearing (NH) subjects (median age: 25 years) and 14 hearing-impaired (HI) subjects with sensorineural hearing-loss (median age: 64.5 years) participated in the study, which was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-KA-04149-g). The HI listeners were recruited *via* a newspaper advertisement. All NH subjects had audiometric thresholds equal to or below 20 dB HL at all tested frequencies. Audiograms of HI subjects are given in Fig. 2.1. Except for subjects 11 and 14, all HI subjects had symmetric or nearly symmetric audiograms and mean low-frequency hearing thresholds below 45 dB HL. Throughout the chapter, “low-frequency hearing thresholds” refer to the average hearing thresholds between 125 and 2000 Hz over both ears. As no individual clinical diagnoses were available, the subjects were asked to report the origin of their hearing loss verbally. These self-reported origins of hearing loss, mean low-frequency hearing thresholds, as well as the gender and age of each subject are listed in Table 3.1.

3.2.2 Experimental set-up

All measurements were carried out *via* a PC in a double-walled sound-attenuating listening booth. Unless otherwise specified, all test procedures were implemented

Table 3.1: From left to right: Subject ID (letters for NH and numbers for HI subjects) • Gender (f for female, m for male) • Age on first day of study • Musical experience (M.): + denotes a musically-trained listener • Self-reported origin of hearing loss • Mean right-ear (R) and left-ear (L) hearing threshold between 125 and 2000 Hz (LF-HL), in dB HL • Right (R) and left (L) tympanogram type • Ipsilateral (I) and contralateral (C) acoustic reflex measured in the right (R) and left (L) ear: + denotes a clear response, ± a weak response, and − an absent response • Overall SNR of CEOAE response in the right (R) and left (L) ear: + denotes a clear response, ± a weak response, and − an absent response • Results of the binaural pitch scale test: Y indicates that the scale was immediately perceived at MCL, NY at MCL+10, NNY at MCL+20, and NNN indicates that only noise was perceived at the three levels • Binaural most comfortable level (BMCL) at 0.5 and 1 kHz, in dB SPL • Binaural (B), right-ear (R), and left-ear (L) L_{10} levels at 0.5 and 1 kHz, in dB SPL • Relative asymmetry of FM detection thresholds (FM-As.) at 0.5 and 1 kHz (in %) • Relative asymmetry of the ERB of auditory filter estimates at 0.5 kHz (in %).

					LF-HL		Tymp.		I-Reflex		C-Reflex		OAE-SNR			BMCL		B-L_{10}		R-L_{10}		L-L_{10}		FM-As.		ERB
#	Sex	Age	M.	Self-reported origin	R	L	R	L	R	L	R	L	R	L	Scale	.5k	1k	.5k	1k	.5k	1k	.5k	1k	.5k	1k	As.
A	m	24.6	+	none	0	2	A	A	+	+	+	+	4.5+	3.1+	Y	77	74	47	45	44	39	45	40	8	2	18
B	m	25.6	+	none	0	-2	A	A	+	+	+	+	3.9+	5.8+	Y	66	65	31	29	41	29	35	25	1	13	15
C	m	21.1	+	none	-4	-6	A	A	+	+	+	+	3.3+	2.0+	Y	58	57	30	30	35	30	33	31	15	17	17
D	f	21.7	−	none	-1	0	A	A	+	+	+	+	3.1+	1.7+	Y	72	85	46	47	53	49	52	48	10	24	1
E	f	47.3	+	none	2	2	A	A	+	+	+	+	2.1+	2.9+	Y	77	63	47	37	46	45	51	45	19	17	15
F	m	26.7	+	none	2	1	A	A	+	+	+	+	2.8+	2.2+	Y	64	54	39	28	39	33	41	31	12	20	13
G	f	24.2	+	none	0	-1	A	A	+	+	+	+	2.9+	2.6+	Y	70	62	45	31							
H	m	25.3	−	none	-1	-2	A	A	+	+	+	+	8.6+	8.7+	Y	66	63	33	31							
1	f	67.6	+	unknown	24	20	A	A	±	+	−	+	1.9−	1.2−	Y	71	76	48	55	52	56	56	51	9	5	2
2	f	46.5	−	innate	18	25	A	A	+	−	+	−	8.4+	8.4+	Y	76	61	46	34	41	40	46	47	61	14	1
3	m	66.3	−	unknown	19	19	A	A	+	+	+	+	3.1+	1.4−	NY	66	67	47	48	50	49	48	49	9	12	11
4	f	64.8	−	unknown	21	24	A	A	+	+	+	+	2.5+	1.8−	NY	73	70	53	47	52	54	59	49	5	8	9
5	m	68.7	−	noise exposure	20	17	A	A	+	+	+	+	1.4+	1.0−	NY	72	75	50	54	54	52	51	59	11	2	6
6	f	66.3	−	unknown	17	19	A	A	+	+	+	+	1.9+	2.7+	Y	76	70	51	46	51	52	54	53	12	16	4
7	m	59.4	−	noise exposure	6	8	A	As	±	±	±	±	1.9±	2.0−	Y	76	76	53	47	56	46	57	45	14	1	2
8	m	65.5	−	unknown	29	27	A	A	+	+	+	+	0.6−	1.2−	NNN	72	71	50	57	46	50	42	51	29	10	9
9	m	55.1	−	high-voltage shock	19	21	A	C	+	+	+	+	3.5+	1.2−	NY	71	75	51	50	52	48	49	49	11	0	22
10	f	63.7	+	hereditary	43	36	A	A	+	±	+	±	1.3−	1.3−	NNN	84	82	64	65	68	72	64	65	24	10	37
11	m	63.3	+	neural disorder	^{1}23	^{1}8	A	A	+	+	+	+	1.0−	1.0−	NY	74	69	51	40	66	60	59	52	19	6	16
12	f	35.6	−	brain trauma	10	9	A	A	+	+	+	+	8.1+	3.6+	Y	71	67	47	43	37	39	39	39	12	8	14
13	f	56.2	+	brain trauma	11	10	A	A	+	+	+	+	3.9+	4.7+	Y	61	62	38	36	46	48	47	47	0	16	20
14	f	66.5	−	brain trauma	^{1}45	^{1}66	A	A	−	±	−	−	0.8−	0.7−	NNN	79	79	67	67	66	64	93	78	9	7	6

---

[a] Asymmetry of 15 dB or more in hearing threshold.

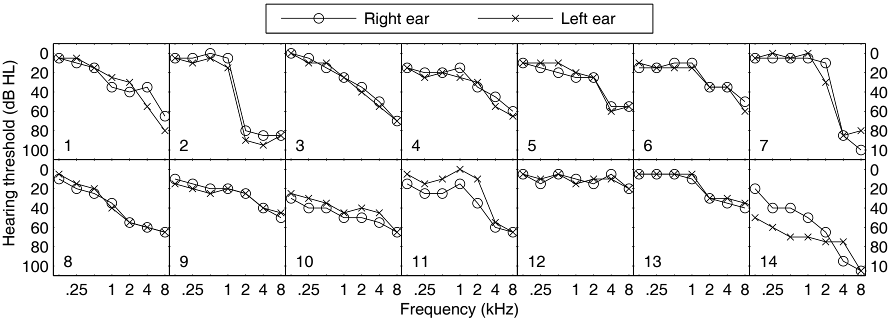


Figure 3.1: Air-conduction pure-tone hearing thresholds of the 14 hearing-impaired subjects. Subject number is indicated at the bottom left of each audiogram.

in MATLAB, the stimuli were presented through Sennheiser HDA200 audiometric headphones connected to an RME DIGI 96/8 24-bit D/A converter, and a 48-kHz sampling frequency was used. Calibration was performed using a B&K 2636 sound level meter, a B&K 4153 artificial ear, and a B&K 4230 artificial-ear calibrator. Additionally, 128-tap linear-phase FIR equalization filters were applied to all broadband stimuli, in order to flatten the headphone frequency response. Testing was divided in experimental sessions of maximum 2 hours, with no more than one session per day per listener. The total testing time was *ca.* 9-10 hours per listener. The audiometric measures were performed first, followed by the loudness perception tests and the binaural pitch experiments. The remaining tests were then conducted in the same order as presented below.

3.3 Binaural pitch experiments

Two pitch-evoking noise stimuli were generated: a binaural pitch (BP) stimulus, and a similar-sounding stimulus evoking a monaurally-detectable pitch (MP). The BP stimulus contained a Huggins' pitch, as this configuration has been reported to evoke the most salient pitch among different types of binaural pitches (Akeroyd *et al.*, 2001; Santurette and Dau, 2007). The pitch in the MP stimulus was created by raising the amplitude of the noise in both ears over a narrow frequency range. The total noise bandwidth was 4 kHz for all stimuli and the transition band of the BP stimulus, over which the interaural phase difference varied linearly from 0 to 2π, had a bandwidth equal to 16% of the band's center frequency.

3.3.1 Scale test

Method

In this short test, the listeners were presented with a sequence of 10 musical notes forming a major scale, ranging from C_5 (523.25 Hz) to C_6 (1046.50 Hz), played with the BP stimulus. The procedure and note frequencies were identical to those described in Santurette and Dau (2007), except that 1-ms ramps were used between intervals, such that the 10-s noise stimulus was perceived as continuous. After one presentation of the stimulus, subjects were asked to verbally report whether something else than noise could be perceived. The test was first performed at an overall level equal to the binaurally-measured most comfortable level (MCL) at 500 Hz (see section 3.5 and Table 3.1). If the test was negative, *i.e.*, only noise was perceived, it was repeated at MCL+10 dB, and if still negative, at MCL+20 dB.

Results

The individual results are summarized in Table 3.1. "Y" indicates that binaural pitch was immediately perceived at MCL, *i.e.*, that the listener reported hearing a pitch within the noise after the first stimulus presentation. "NY" indicates that a pitch was first heard during the second presentation (MCL+10), and "NNN" that no pitch was heard in any of the three presentations, *i.e.*, the listener reported hearing noise only, even at MCL+20. All NH listeners and 6 HI listeners could hear binaural pitch immediately at MCL, while 5 other HI listeners could first perceive it at MCL+10. The three remaining HI listeners (8, 10, and 14) only perceived noise at all presentation levels, suggesting impaired binaural-pitch perception.

3.3.2 Pitch contour detection and identification

Method

The ability of the listeners to detect and identify pitch contours generated with BP and MP stimuli were measured and compared, using a procedure similar to that described in Santurette *et al.* (2010). In each trial, a sequence of five musical notes (Table 4.1(a)), corresponding to frequencies within the range of strongest salience of Huggins' pitch (Santurette and Dau, 2007), were presented to the listeners, such that they formed one of the five possible pitch contours listed in Table 4.1(b). The task of the subjects was to press one of six buttons on a computer screen after each stimulus presentation. Five buttons corresponded to the five possible pitch contours, represented by symbols, while the words "no melody" appeared on the sixth button. The listeners were instructed to press the "no melody" button if no pitch was heard at all. If any pitch was heard, they were asked to press the symbol corresponding to the perceived pitch contour. The "no melody" option was included so that both detection and contour recognition could be tested within a single experiment.

The BP and MP stimuli were generated and adjusted for equivalent pitch salience as described in Santurette *et al.* (2010). In that study, following a salience-adjustment experiment in NH listeners, a linear relationship[1] was derived between the broadband-noise level and the level of the additional narrow-band noise in MP producing an equally-salient pitch to that of the corresponding BP stimulus. The same linear relationship was used here, with the level of the broadband noise individually adjusted to the binaurally-measured MCL at 500 Hz (Table 3.1). 30-ms onset and offset ramps were applied to the overall stimulus. Three different note durations were used (300, 600, and 900 ms). Subjects were presented 30 trials for each combination of stimulus type and note duration. Each possible pitch contour was presented an equal number of times. In addition to these 180 trials containing a pitch contour, 36 trials containing no pitch contour (diotic white noise only) were presented, one third of them corresponding to each duration. This made it possible to evaluate false

---

[1] $L_{\mathrm{NBN}} = 1.07 \times L_{\mathrm{BBN}} - 15.69$, where L_{BBN} is the overall level of the broadband noise in both BP and MP stimuli and L_{NBN} is the overall level of the additional narrow band of noise in the MP stimulus.

alarms and to avoid the possibility of subjects never pressing the "no melody" button. The experiment was divided into two blocks and trials were presented in a random order within each block. Before these two experimental blocks, each subject was first introduced to the different pitch contours played with pure-tone stimuli. In addition, at least one short 18-trial practice block was performed with pure tones to ensure that the task was correctly understood. Subjects were not informed about the existence of different stimulus types.

Table 3.2: Note frequencies and pitch contours used in the pitch contour identification experiment. (a) Note frequencies. (b) Pitch contours.

(a) Note frequencies

Note	Frequency
C5	523.25 Hz
D5	587.32 Hz
E5	659.26 Hz
F5	698.46 Hz
G5	783.99 Hz

(b) Pitch contours

Contour	Note sequence
Rising only	C5-D5-E5-F5-G5
Falling only	G5-F5-E5-D5-C5
Rising then falling	C5-D5-E5-D5-C5
Falling then rising	G5-F5-E5-F5-G5
Constant	E5-E5-E5-E5-E5

Results and discussion

All NH listeners and 11 HI listeners obtained overall detection scores above 95% for both MP and BP stimuli, indicating a clear pitch sensation, independent of the use of binaural or monaural cues. HI subject 11 was also able to perceive both pitch types clearly (MP: 86%, BP: 81%). The two remaining HI listeners could perceive MP in 97% of trials, but failed to perceive BP stimuli (subject 10: 1%, subject 14: 0%). Despite his negative response in the scale test, subject 8 was clearly able to detect BP as well as MP. This suggests that a negative response in the scale test is not sufficient to conclude that binaural-pitch perception is absent, which was also observed by Nitschmann *et al.* (2010) in one of their subjects. In summary, for HI listeners, the BP stimulus was either as easy to detect as the MP stimulus, or not perceived at all despite accurate MP detection. This rules out a general difficulty in extracting tonal objects from background noise in HI listeners unable to perceive binaural pitch, and confirms that the lack of accurate binaural information must be a

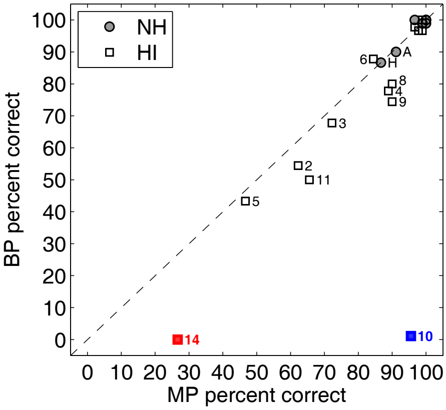


Figure 3.2: Pitch contour identification scores for the MP (horizontal axis) and BP (vertical axis) stimuli, expressed as the percentage of trials containing a pitch contour in which the pitch contour was correctly identified.

crucial factor in the observed deficit. Stimulus duration was found to have no effect on the detection scores of HI listeners (MP: p=0.349, BP: p=0.379, Kruskal-Wallis test). Low false alarm rates, *i.e.*, proportions of trials containing no pitch contour in which another button than "no melody" was pressed, were obtained in most subjects (mean 2.8%, median 1.0%). Individual false alarm rates higher than 20% were obtained in subject A (36%), subject 2 (36%), and subject 7 (22%).

The ability of the listeners to correctly identify the pitch contours is represented in Fig. 3.2. The proportions of trials containing a pitch contour in which the pitch contour was correctly identified are plotted for the MP and BP stimuli against each other. Here and in all following figures, the relevant individual results are indicated using letters for NH subjects and numbers for HI subjects (see Table 3.1). NH listeners all obtained identification scores above 85% and similar results for the MP and BP stimuli. A larger variability was found in the HI group, with 5 listeners showing similar performance to the NH group (subjects 1, 6, 7, 12, and 13), and 7 other listeners obtaining scores below 80% either for the BP stimulus only or for both stimulus types. For the latter group of subjects, BP identification was always poorer than MP identification (points below the diagonal line in Fig. 3.2), despite similar detection scores for MP and BP.

This difference was significant (p=0.017, Wilcoxon paired signed-rank test, α=5% here and throughout the chapter), indicating a reduced salience or musicality of BP compared to MP for these listeners. The two HI subjects who could not perceive binaural pitch performed very differently in identifying MP contours: Subject 14 showed difficulty in pitch contour identification, while subject 10 did not. No effect of stimulus duration was found on pitch contour identification in the HI group (MP: p=0.458, BP: p=0.726, Kruskal-Wallis test).

Overall pitch identification scores of the HI listeners (MP score for subjects 10 and 14, mean MP and BP score for all other subjects) were significantly correlated to pitch contour identification with pure-tone stimuli obtained in the practice blocks of the experiment (p=0.002, ρ=0.75). This suggests that the low salience of MP and BP, in comparison to that of pure tones, was not the main factor responsible for reduced pitch contour identification in some HI listeners. Only for subjects 5 and 14 were pure-tone scores much higher than MP and BP scores, which suggests a detrimential effect of using noise-based stimuli for these two specific subjects. The pitch contour identification abilities of the listeners were overall not found to rely on their musical experience (column “M” in Table 3.1).

In addition to binaural pitch detectability, two additional outcomes of this test were used to perform a correlation analysis with the different measures of the auditory profile described further below: average pitch contour identification scores, as well as the difference in performance between MP and BP identification scores as an estimate of the binaural disadvantage for pitch contour identification. In the following, “MP+BP scores” thus refer to the average pitch identification scores of HI listeners with MP and BP stimuli (except for subjects 10 and 14, for whom it refers to performance with the MP simulus only). For listeners who could perceive binaural pitch, the “MP−BP score” refers to the difference between the MP and BP scores divided by the MP+BP score. All correlation coefficients mentioned in this chapter correspond to Spearman’s ρ and, unless specified otherwise, are calculated for the HI group. No correction was applied to the reported p-values[2], which were calculated from the exact permutation distributions.

In addition to the correlations of MP+BP and MP−BP scores with the different outcomes, the performance of the two subjects who could not perceive binaural pitch at all (10 and 14) in the different auditory profile measures will be in focus in sections IV to IX. The results of subjects 9 and 11 may also be of particular interest, because these two listeners obtained the highest MP−BP scores in the HI group (data points furthest away from the diagonal line in Fig. 3.2).

The detailed methods of the auditory profile tests presented in the following can be found in the appendix of this chapter.

3.4 Audiometric measures

3.4.1 Pure-tone audiometry

Pure-tone air-conduction audiograms of all HI subjects are given in Fig. 3.1. None of the listeners showed an air-bone gap, indicating no sign of conductive dysfunction. All hearing losses were thus of the sensorineural type. Subject 12 showed no hearing-threshold elevation at any audiometric frequency, and is thus to be classified as suffering from an obscure dysfunction (*e.g.*, Saunders and Haggard, 1989; King and Stephens, 1992; Strelcyk and Dau, 2009).

It is worth noting that the two listeners who could not perceive BP (10 and 14) were also the ones with the highest average low-frequency hearing thresholds (LF-HL in Table 3.1). This raises the question of the influence of sensation level on BP detection scores. However, the use of the MCL as a testing level should have ensured

---

[2] A total of 33 outcome measures were compared in this study, raising the issue of multiple testing. Among all possible comparisons, 105 correlation coefficients were used to interpret the relationship between the different outcomes. In order to find an appropriate significance level for this large set of correlations, a p-value plot (Schweder and Spjøtvoll, 1982) was drawn, and the marginal significance of the largest coefficients was calculated using the asymptotic theory of Buckley and Eagleson (1986). According to these two methods, a significance level α=5% was reasonable in the present context. Therefore, no correction was applied to the p-values reported in this chapter.

sufficient audibility in the binaural pitch tests, and neither subject 10 nor 14 benefited from a raised sensation level in the scale test. Moreover, in the study of Nitschmann *et al.* (2010), some listeners with moderate low-frequency hearing loss, similar to that of subject 10 in the present study, were able to perceive binaural pitch. Therefore, it is unlikely that audibility alone was responsible for the absence of binaural pitch percept found in subjects 10 and 14. Neither MP+BP scores nor MP−BP scores were correlated with mean hearing thresholds at 0.5 and 1 kHz (MP+BP: p=0.234, MP−BP: p=0.499).

3.4.2 Tympanometry and acoustic reflex

Tympanograms as well as ipsilateral and contralateral stapedius reflex curves were obtained in both ears for each subject. The obtained tympanogram types, as described in Gelfand (2001a) using the classification of Jerger (1970), are given in Table 3.1. All subjects obtained type A tympanograms in both ears, except subject 7 who showed a shallow admittance peak in the left ear (type As), and subject 9 who showed a negative-pressure peak in the left ear (type C). Together with the absence of an air-bone gap, this confirms the sensorineural nature of the hearing loss in all listeners. Although type C tympanograms are usually associated with Eustachian tube dysfunction (Feldman, 1977; Gelfand, 2001a), the symmetry in the audiogram of subject 9 suggests that this did not affect his hearing ability.

The nature of the individual stapedius reflex responses are given in Table 3.1 for ipsilateral (I-Reflex) and contralateral (C-Reflex) stimulation. A “+” sign indicates the presence of a clear response at a presentation level of 100 dB SPL or below for at least one of the three test frequencies (0.5, 1, and 2 kHz). A “−” sign indicates that no response was obtained at 100 dB SPL for any test frequency. A “±” sign indicates a weak or undetermined response. Most HI listeners obtained patterns of responses which did neither suggest a conductive dysfunction nor a deficit along the neural sensory pathways. Only subject 2 obtained a clear pattern usually associated with VIII$^{\text{th}}$ nerve disorder on the left side (Gelfand, 2001a). The absence of a clear response for subject 14 may reflect her elevated hearing threshold compared to other subjects, as auditory reflex thresholds above 100 dB SPL are not uncommon in cases of moderate to moderately-severe sensorineural hearing loss (Gelfand, 2001a).

3.4.3 Otoacoustic emissions

Click-evoked otoacoustic emissions (CEOAEs) were measured in both ears for each subject. The frequency spectrum of the CEOAE response and the noise-floor spectrum were plotted and compared for each measured ear. An artefact-rejection template was applied to remove ca. 10% of the noisiest epochs, in order to increase the signal-to-noise ratio (SNR), before the click responses were averaged. The obtained overall SNRs are given in Table 3.1, together with a visual assessment of the strength of the response at low frequencies (500 to 1500 Hz). A "+" sign indicates a clearly visible response, a "±" sign a weak or undetermined response, and a "−" sign a response which was not distinguishable from the noise floor. While OAEs were present in all NH ears, only HI listeners 2, 6, 12, and 13 showed a clearly visible response in both ears, suggesting the presence of normal OHC function over at least a large part of the low-frequency range. This is consistent with the hearing thresholds of these subjects lying within the NH range up to at least 1 kHz, as CEOAEs are present in about 98% of NH ears and in more than 95% of HI ears with sensorineural hearing loss where the hearing threshold lies below 18 dB HL (Probst *et al.*, 1991). The two listeners with absent binaural pitch perception showed no OAE response, which was expected as their hearing thresholds exceeded 35 dB HL (Probst *et al.*, 1991). The OAE SNRs in the HI group were significantly correlated with the low-frequency hearing thresholds in the corresponding ears (p=0.008, ρ=-0.49), and so were the mean SNRs with the age of the HI listeners (p=0.006, ρ=-0.70).

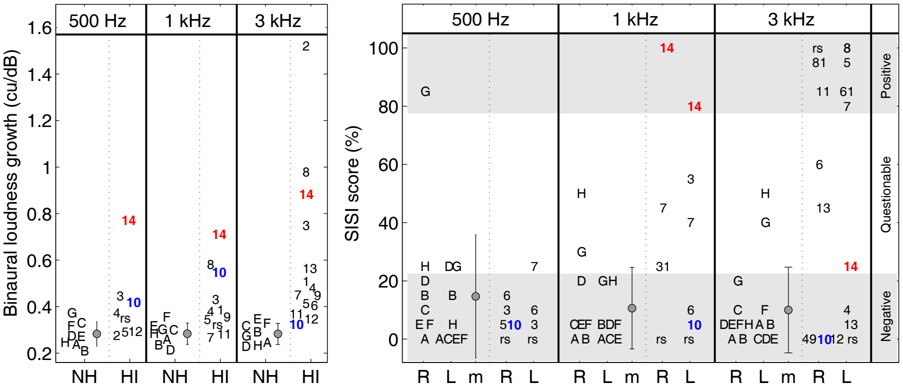


Figure 3.3: Slope of the lower section of the binaural loudness curve at 0.5, 1, and 3 kHz (left panel) and SISI scores (right panel) for NH and HI subjects. Individual results for the right (R) and left (L) ear, and mean and standard deviation across NH subjects (m). “rs” stands for “remaining subjects”.

3.5 Loudness perception

3.5.1 Loudness scaling

An adaptive categorical loudness scaling procedure (*e.g.*, Brand and Hohmann, 2002; ISO 16832, 2006) was used to determine the loudness function and most-comfortable level (MCL). Measurements were performed binaurally at 500, 1000, and 3000 Hz, and monaurally at 500 and 1000 Hz. The MCL in a given condition was defined as the level corresponding to 20 loudness categorical units (cu). The loudness scaling experiment was mainly used as a tool to determine adequate presentation levels in individual HI listeners. These are provided in Table 3.1 for test frequencies of 0.5 and 1 kHz, in the form of the binaurally-measured MCL (BMCL), and in the form of the levels corresponding to 10 loudness cu (L_{10}) for binaural as well as monaural left and right presentations.

In addition to ensuring sufficient audibility of the stimuli in the different experiments, loudness curves may also be used to evaluate the degree of loudness recruitment in each listener. This may help shed light on the location of the individual hearing losses, because loudness recruitment has been associated with a cochlear site of lesion in sensorineural hearing loss (Dix *et al.*, 1948). However, much caution is needed as recruitment is also found in cases of neural auditory disorders (Priede and Coles, 1974) and has been proven to be a weak predictor as a site-of-lesion test (Hood, 1969). The loudness growth at threshold was estimated as the slope, in cu/dB, of the lower section of the fitted loudness curve for each measurement. The obtained values for the binaural measurements at 0.5, 1, and 3 kHz are plotted in the left panel of Fig. 3.3. For readability reasons, ‘rs’ (“remaining subjects”) is used to indicate the mean result of HI listeners not mentioned by their number, as these listeners obtained very similar values. As the left and right loudness-growth values at 0.5 and 1 kHz were found to show a high degree of symmetry and follow the same trend as the binaurally-measured values, only the binaural loudness growth is shown in Fig. 3.3.

At 0.5 and 1 kHz, most HI listeners had normal or near-normal loudness growth, which may be accounted for by the fact that most of them only had elevated hearing thresholds at high frequencies. Reflecting this, the binaural loudness growth values in the HI group were significantly correlated with the average hearing thresholds over both ears for all three test frequencies (500 Hz: p=0.007, ρ=0.68; 1 kHz: p<0.001, ρ=0.97; 3 kHz: p=0.038, ρ=0.56). This is consistent with the findings of Al-Salim *et al.* (2009), who reported loudness growth values that were correlated with hearing thresholds. While these loudness growth estimates may thereby not reveal much about the site of lesion for most listeners, the difference in loudness growth for subjects 10 and 14 is worth mentioning. While subject 14 showed clear recruitment, loudness growth was substantially lower for subject 10, who obtained a value similar to that of subject 12 at 3 kHz, despite a difference of about 40 dB in hearing threshold between these two listeners. According to this, it is likely that a cochlear impairment is present in subject 14, but not in subject 10.

3.5.2 Short Increment Sensitivity Index

A Short Increment Sensitivity Index (SISI) test (Jerger *et al.*, 1959) was performed in the left and right ear at 500, 1000, and 3000 Hz. The results may be useful to further assess the presence of loudness recruitment. The right panel of Fig. 3.3

shows the SISI scores obtained by the individual listeners in each ear for the three test frequencies. The listeners indicated by ‘rs’ all obtained the same score. The test results are classified as positive ($\geq$80%), questionable (25-75%), or negative ($\leq$20%), following the stricter of several possible criteria (Buus *et al.*, 1982). While a positive result is typical in cases of cochlear impairment, a negative result is found in most normal-hearing listeners and otherwise usually associated with retrocochlear disorders (Gelfand, 2001b). However, one should keep in mind that SISI scores have a high predictive value for cochlear losses (91%), but only moderate predictive value for retrocochlear losses (48%) (Buus *et al.*, 1982).

The SISI scores of subjects 10 and 14 confirm the discrepancy in their loudness growth estimates. Subject 14 obtained positive or questionable results, again suggesting a cochlear impairment, while subject 10 consistently obtained negative SISI scores at all test frequencies, which might indicate a retrocochlear hearing loss. As other listeners showed little loss of audibility at low frequencies, only their SISI scores at 3 kHz may be informative. Positive scores in both ears were found in subjects 1, 5, 6, 7, and 8, suggesting cochlear hearing loss. Negative scores in both ears were found in subjects 4, 9, and 12. As subject 12 suffered from obscure dysfunction and showed a very strong OAE response, this indicates normal OHC function. The remaining listeners showed asymmetric SISI scores. Subjects 2, 3, and 11 obtained a positive result in the right ear, and a negative result in the left ear. For subject 2, this is consistent with an asymmetry in the acoustic reflex pattern, and with a strong OAE response. For subject 11, the asymmetry might reflect a similar asymmetry in hearing thresholds. Finally, subject 13 obtained a questionable score in the right ear and a negative score in the left ear. Taken together with her strong OAE response, this suggests normal OHC function. Overall, the combined results from the different audiometric tests do not allow further conclusions about the site of impairment for each HI listener.

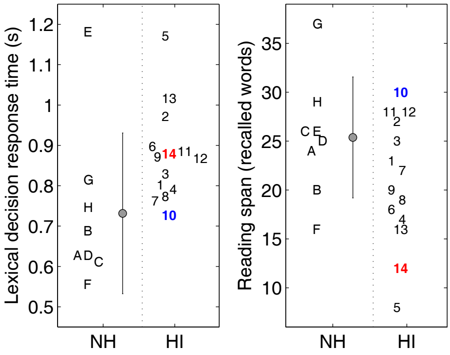


Figure 3.4: Response times in the lexical decision test (left panel) and reading span scores (right panel) for individual NH and HI subjects. The means and standard deviations across NH subjects are also given.

3.6 Cognitive abilities

The ability of the listeners to detect tonal targets and process pitch sequences in background noise may be influenced by high-level factors linked to global processing speed of sensory stimuli, working memory processing and capacity, and decision making, as suggested by studies investigating binaural pitch perception in dyslexic listeners (Chait *et al.*, 2007; Santurette *et al.*, 2010). Moreover, there is evidence that reduced processing speed and working-memory deficits play a role in the difficulty of HI listeners to understand speech in adverse conditions (*e.g.*, van Rooij *et al.*, 1989; van Rooij and Plomp, 1990; Lunner, 2003; Foo *et al.*, 2007). Therefore, these cognitive functions were evaluated in a lexical decision task (*e.g.*, Baddeley *et al.*, 1985) and a reading span test (*e.g.*, Daneman and Carpenter, 1980; Lunner, 2003), in order to investigate a potential effect on performance in the binaural-pitch and auditory-profile tests.

3.6.1 Lexical access and decision making

A lexical decision task similar to that used by Baddeley *et al.* (1985) was performed. The subjects' task was to evaluate words as real or non-existing words. The proportion of correct responses and response times were measured. The percentage of words correctly identified as real or non-existing was above 92% for all NH and HI subjects. The response times of the individual subjects are given in the left panel of Fig. 3.4. The difference in response times between the NH and HI groups was only borderline significant (p=0.045, two-sample t-test), and most HI subjects lied within one standard deviation of the NH mean, including subjects 10 and 14. No correlation was found between response times and either MP+BP scores (p=0.122) or MP$-$BP scores (p=0.688).

3.6.2 Working memory processing and capacity

A reading span test similar to that used by Lunner (2003) was performed. The subjects' task was to evaluate sentences as normal or absurd, and to recall the first or last word in each sentence. The reading span was defined as the total number of correctly recalled words. The reading span scores are given in the right panel of Fig. 3.4. A large variability was found in both subject groups, and the range of reading span scores was similar to that obtained by Lunner (2003). There was no group difference between NH and HI subjects (p=0.130, two-sample t-test). While subject 14 obtained a low reading span, subject 10 obtained the highest score among the HI subjects, indicating no influence of working memory on binaural pitch perception. Moreover, there was no correlation between reading span scores and either MP+BP scores (p=0.456) or MP$-$BP scores (p=0.688).

3.6.3 Discussion

Overall, cognitive function, as measured by these two tests, was neither related to the ability to perceive binaural pitch, nor to the difference in identification scores between MP and BP stimuli. As a whole, performance in the cognitive tests was also not correlated with the ability of the listeners to perform the pitch contour identification task. However, subject 5 performed substantially worse than all other HI subjects in both cognitive tasks, and subject 14 had a reading span below 15 words. As these two subjects were also those with the lowest MP+BP scores, an influence of reduced cognitive abilities on pitch contour identification cannot be completely excluded for

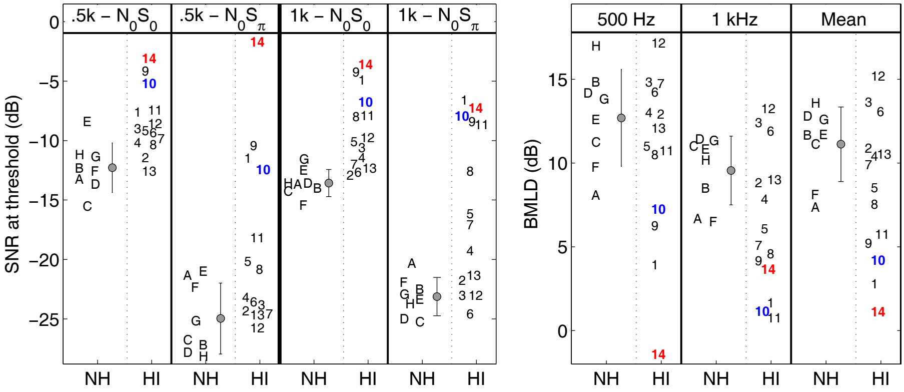


Figure 3.5: Binaural masked thresholds at 0.5 and 1 kHz in N_0S_0 and N_0S_π conditions (left panel) and resulting binaural masking level differences (right panel) for individual NH and HI subjects. The means and standard deviations across NH subjects are also given, as well as the mean BMLD across the two test frequencies.

these listeners. Interestingly, these were also the two listeners for whom pitch contour identification was affected by the use of noise-based stimuli. Finally, the fact that subject 10 was the best performer in both cognitive tests clearly indicates that the absence of binaural pitch percept does not imply reduced cognitive function.

3.7 Binaural masking release

3.7.1 Binaural Masking Level Difference (BMLD)

The masking thresholds of 500-Hz and 1000-Hz tones in background noise were measured binaurally in two conditions. In the first condition, both the signal and the noise were diotic (N_0S_0 condition). In the second condition, the noise was diotic and an interaural phase shift of 180° was introduced in the signal (N_0S_π condition). For a given tone frequency, the BMLD was calculated as the difference in threshold between the N_0S_0 and the N_0S_π condition.

Fig. 3.5 shows the masked thresholds obtained at 500 Hz and 1 kHz for the N_0S_0 and

N_0S_π conditions (left panel), as well as the resulting BMLDs at 500 Hz and 1 kHz and the mean values for the two frequencies (right panel). With the exception of subject 14 at 500 Hz, masked thresholds were always lower in the N_0S_π than in the N_0S_0 conditions, indicating a release from masking for all listeners with the dichotic tone. The SNRs at threshold were significantly higher in the HI group than in the NH group for all conditions (500 Hz–N_0S_0: p=0.002, 500 Hz–N_0S_π: p=0.037, 1 kHz–N_0S_0: p=0.001, 1 kHz–N_0S_π: p=0.006, two-sample t-tests). However, the group difference was not significant for the resulting BMLDs (500 Hz: p=0.276, 1 kHz: p=0.066, Mean: p=0.120, two-sample t-tests), and most HI listeners obtained BMLDs within the NH range. This is consistent with earlier reports of BMLDs (Staffel *et al.*, 1990; Gabriel *et al.*, 1992; Strelcyk and Dau, 2009), in which HI listeners showed elevated thresholds in both the diotic and dichotic conditions, leading to less pronounced differences with NH listeners in terms of masking release.

The two listeners with absent binaural pitch perception (10 and 14) obtained both elevated masked thresholds and a largely reduced masking release. Interestingly, the two subjects with the largest MP−BP scores (9 and 11) were also among the listeners with both the highest masked thresholds and the lowest BMLDs. This indicates that impaired binaural pitch perception coexists with a loss of binaural advantage in background noise. However, the latter does not imply the former (*cf.* subject 1). Subject 12 obtained the highest BMLDs among all listeners, including those from the NH group, confirming the findings of Strelcyk and Dau (2009) that listeners with obscure dysfunction do not show a deficit in binaural masking release.

Overall, there was no correlation in the whole HI group between mean BMLDs and either MP−BP scores (p=0.307) or MP+BP scores (p=0.445). The mean BMLD was found to be significantly correlated with the asymmetry in hearing threshold (p=0.013, ρ=-0.65), suggesting that using the same presentation level in both ears might affect binaural masking release in cases of asymmetric hearing loss. BMLDs were also significantly correlated with mean OAE SNRs (p=0.006, ρ=0.71), as well as hearing thresholds at 500 Hz (p=0.003, ρ=-0.73), but not at 1 kHz (p=0.258). Finally, the correlation of BMLDs with dichotic masked thresholds was highly significant

(500 Hz: $p<0.001$, 1 kHz: $p<0.001$), indicating that performance in the N_0S_π condition is sufficient to predict the amount of masking release.

3.7.2 Binaural Intelligibility Level Difference (BILD)

The subjects' SRT was measured in five different conditions, using Danish closed-set sentences (DANTALE II, Wagener *et al.*, 2003). In the reference condition, both the speech signal and the masking noise were located in front of the listener and presented binaurally (bS_0N_0). In all other conditions, the target speech was kept in front of the listener, while the noise interferer was located on one side, with an azimuthal angle of 105°, where the largest amount of spatial release from masking is obtained (Peissig and Kollmeier, 1997). In two of the conditions, the speech and noise were presented binaurally, with the noise located either on the right (bS_0N_{105}) or the left (bS_0N_{255}) side of the head. In the remaining two conditions, the speech and noise were presented monaurally to the ear opposite to the noise location, by setting the sound card attenuation to infinity in the right (mS_0N_{105}) or left (mS_0N_{255}) channel. The intelligibility level difference (ILD) was defined as the total amount of spatial release when the noise interferer was moved to the side:

- $\mathrm{ILD_{right}} = \mathrm{SRT}(bS_0N_0) - \mathrm{SRT}(bS_0N_{105})$;
- $\mathrm{ILD_{left}} = \mathrm{SRT}(bS_0N_0) - \mathrm{SRT}(bS_0N_{255})$.

The binaural intelligibility level difference (BILD) was defined as the contribution of binaural interaction to the spatial release, *i.e.*, the amount of spatial release not due to better-ear listening. It can be expressed as the difference in spatial release (or SRT) between the binaural and monaural conditions:

- $\mathrm{BILD_{right}} = \mathrm{SRT}(bS_0N_{105}) - \mathrm{SRT}(mS_0N_{105})$;
- $\mathrm{BILD_{left}} = \mathrm{SRT}(bS_0N_{255}) - \mathrm{SRT}(mS_0N_{255})$.

In the left panel of Fig. 3.6, the reference SRTs (obtained binaurally with the interfering noise at 0° azimuth) are given for each individual subject. In the right

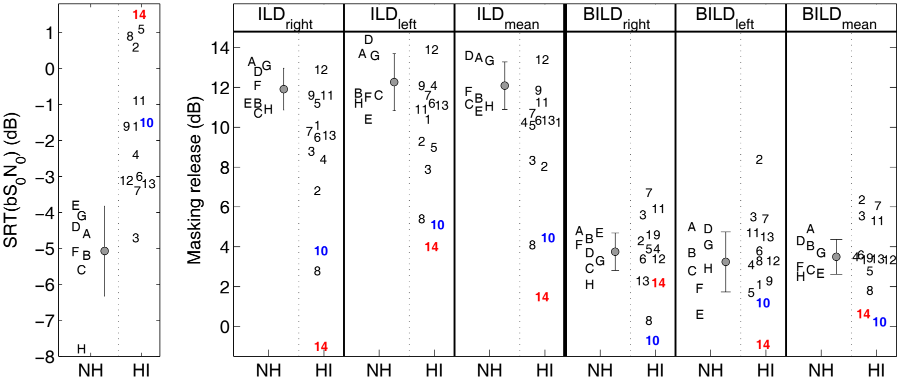


Figure 3.6: Speech reception thresholds in the bN_0S_0 condition (left panel) and binaural intelligibility level differences (right panel) for individual NH and HI subjects. In the right panel, both the total amount of masking release (ILD) and the estimated binaural contribution (BILD) are given for the right and left ear. The means and standard deviations across NH subjects are also given, as well as the mean ILD and BILD across ears.

panel, the total drop in SRT can be seen when the noise is moved to the right ($\text{ILD}_{\text{right}}$) or to the left ($\text{ILD}_{\text{left}}$), and is also represented as the average spatial release over these two conditions. The respective contributions of binaural processing to this spatial release (BILD) are then given.

The reference SRTs provide a measure of speech intelligibility in noise. All HI listeners except subject 3 obtained elevated SRTs compared to NH listeners, and the difference between the two groups was significant ($p<0.001$, two-sample t-test). A significant correlation was found between SRTs and MP+BP scores ($p=0.009$, $\rho=-0.67$). As SRTs were not correlated with pure-tone pitch identification scores ($p=0.415$), this reflects the difficulty of some HI listeners in identifying signals in the presence of background noise, whether these signals are words or musical melodies. Significant correlations were also found between SRTs and N_0S_π masked thresholds (500 Hz: $p=0.014$, $\rho=0.64$; 1 kHz: $p=0.035$, $\rho=0.57$), and thereby BMLDs (500 Hz: $p=0.006$, $\rho=-0.70$; 1 kHz: $p=0.037$, $\rho=-0.56$). Cognitive abilities have been shown to play an important role for speech reception in background noise (see Akeroyd (2008) for a review), and especially the reading span of HI listeners was found to correlate with measures of speech intelligibility in noise (*e.g.*, Lunner, 2003; Foo *et al.*, 2007).

However, neither the lexical-decision response times (p=0.441), nor the reading span scores (p=0.463, p=0.629 with age and audibility controlled for), were correlated with the reference SRTs in the present study. No correlation was found between SRTs and MP−BP scores (p=0.102), and there was a correlation of SRTs with low-frequency hearing thresholds (p=0.021, ρ=0.61). Subject 12 obtained an elevated SRT compared to all NH listeners, despite a normal audiogram.

The total amount of spatial release when the interfering noise was moved to the side (ILD) was found to be significantly higher in NH than in HI listeners (ILD_{mean}: p=0.019, two-sample t-test), which is consistent with the findings of Peissig and Kollmeier (1997). However, in the present study, a significant correlation was found between mean ILDs and low-frequency hearing thresholds ($p<0.001$, ρ=-0.79). Another significant, although borderline, correlation was found between mean ILDs and reference SRTs (p=0.046, ρ=-0.54). Despite the significant group difference, most HI listeners obtained ILD_{mean} values similar to those of NH listeners. Subjects 2 and 3 showed a slight decrease in ILD, while subjects 8, 10, and 14 had a considerably reduced spatial release from masking.

Binaural processing was found to account for about 29% of the total spatial release in NH listeners, with a mean BILD of 3.5 dB. This is a slightly lower binaural advantage than the ones reported in the review of Blauert (1997) and in more recent studies (Johansson and Arlinger, 2002; Goverts and Houtgast, 2010). However, most of these studies used experimental designs in which SRTs were compared for homophasic *vs.* antiphasic speech or noise, and the use of a spatial design with head-related transfer functions (HRTFs) may be responsible for the different outcome in the present study. Overall, no group difference was found between NH and HI listeners for the mean BILD (p=0.961, two-sample t-test). This indicates that all HI listeners who obtained mean ILDs above 8 dB showed some advantageous contribution of binaural processing to the spatial release. The two listeners who could not hear binaural pitch showed no binaural advantage in spatially segregating speech from noise. Mean BILDs were not correlated with MP+BP scores (p=0.929), MP−BP scores (p=0.720), or low-frequency hearing thresholds (p=0.053). A significant correlation was found between mean BILDs and mean BMLDs (p=0.045, ρ=0.54), as well as with BMLDs at 500 Hz (p=0.027, ρ=0.59) but not at 1 kHz (p=0.205).

3.8 Temporal fine structure processing

The ability of the listeners to use TFS cues was evaluated binaurally in an IPD detection task, in which the upper frequency limit for detecting a 180° phase shift in a tone was measured. The carrier frequency was chosen as the tracking variable in order to investigate the effect of sensorineural hearing loss on the upper frequency limit of binaural phase-locking. Moreover, as the maximum IPD in the BP stimulus was 180°, the inability of a listener to detect such a large IPD within the most salient range of BP would provide a clear explanation for the absence of binaural pitch sensation. As the IPD detection task may rely on both peripheral TFS processing and the integration and processing of binaural information, frequency-modulation detection thresholds (FMDTs) at a 2-Hz FM-rate were additionally used as a monaural measure. This is because FM detection at low FM-rates is thought to primarily rely on accurate TFS processing (Moore and Sęk, 1996; Lacher-Fougère and Demany, 1998; Strelcyk and Dau, 2009). Only six of the NH subjects were available to participate in these experiments.

3.8.1 Interaural phase difference detection

The upper frequency threshold for detectability of a 180° interaural phase difference (IPD) was measured. The results are given in the left panel of Fig. 3.7. NH listeners were able to detect the IPD up to a carrier-frequency of 1336 Hz on average. This value and the range of obtained thresholds are in line with the results of Ross *et al.* (2007b), obtained *via* a similar behavioral method as well as cortical auditory evoked magnetic responses to IPD changes. The fact that subject E's threshold (899 Hz) was lower than those of all other NH listeners, who all lay above 1250 Hz, may reflect her age difference with the rest of the NH group. This is consistent with the significant

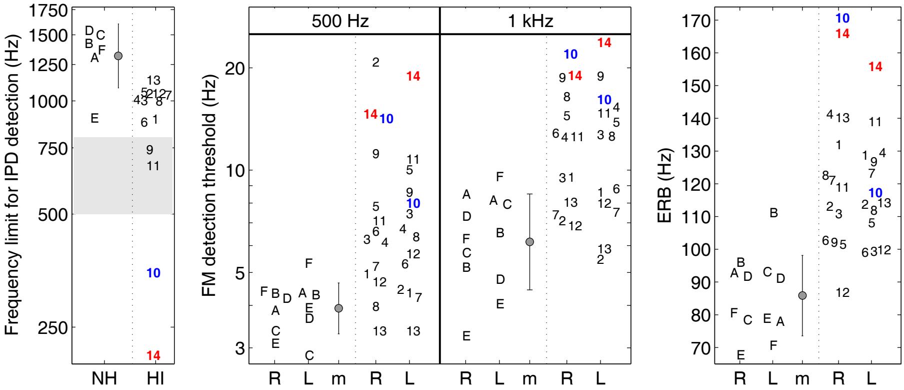


Figure 3.7: Upper frequency limit for detection of a 180° interaural phase difference (left panel), frequency modulation detection thresholds at 0.5 and 1 kHz (center panel), and ERB of the auditory filter estimates at 500 Hz (right panel) for NH and HI subjects. Individual results for the right (R) and left (L) ear, and mean and standard deviation across NH ears (m). In the left panel, the gray area indicates the frequency range of the notes used in the pitch contour identification experiment.

decrease in thresholds found by Ross *et al.* (2007a) between young and middle-aged subjects.

A significant group difference was found between NH and HI listeners (p=0.026, two-sample t-test). Despite this, most HI listeners' thresholds were found to lie around 1000 Hz. For such listeners, there is thus no apparent deficit in detection of a 180° IPD, given the age difference between the NH and HI groups. Subjects 10 and 14, however, obtained dramatically lower thresholds than all other HI listeners (349 and 210 Hz, respectively). Therefore, there is a specific deficit for these two listeners in using interaural phase cues, even in quiet, which cannot be accounted for by an age factor. Furthermore, thresholds in the HI group were not correlated with age (p=0.368). As binaural pitch perception relies on the introduction of an IPD in a noise stimulus, and as the note frequencies used in the binaural pitch experiments were all above 500 Hz, the inability of subjects 10 and 14 to detect IPDs at such frequencies explains why they could not perceive any pitch in the BP stimulus. Interestingly, subjects 9 and 11, who had the highest MP$-$BP scores, obtained thresholds within the frequency range of the notes used in the pitch contour identification experiment (gray area in the left panel of Fig. 3.7). This suggests that they may not have heard the higher notes when played with the BP stimulus, leading to higher MP−BP scores.

IPD detection thresholds were found to be significantly correlated with BMLDs at 500 Hz (p=0.025, ρ=0.59) and 1 kHz (p=0.009, ρ=0.67), consistent with the fact that both tasks involve the use of accurate interaural timing representations for detecting either the presence of a tone in noise (BMLD) or a change in the spatiality of a sound image (IPD detection). Following the correlation of BMLDs with OAE SNRs, IPD detection thresholds were also correlated with the mean OAE SNRs of the listeners (p=0.050, ρ=0.53). No correlation was found between IPD detection thresholds and either MP+BP (p=0.315) or MP−BP (p=0.508) scores. The correlation of IPD detection thresholds with low-frequency hearing thresholds was borderline significant (p=0.046, ρ=-0.54). This raises the questions of whether the observed deficits are suprathreshold deficits or a direct consequence of a loss of audibility, and whether there was an effect of using different sensation levels across subjects on the obtained thresholds. Using IPD detection tasks in which the IPD was the tracking variable, Lacher-Fougère and Demany (2005) found no effect of sensation level on performance, and Strelcyk and Dau (2009) obtained thresholds that were not correlated with audibility. These findings, together with the fact that stimuli were adjusted for equal loudness and sensation levels never fell below 20 dB SL in the present study, strongly suggest that factors other than audibility were responsible for the observed deficits in binaural TFS processing.

3.8.2 Frequency modulation detection

Pure-tone FMDTs were measured in quiet for the subjects' left and right ear at 500 and 1000 Hz. The obtained FMDTs are shown in the center panel of Fig. 3.7. The mean FMDTs over NH ears were 3.97 Hz (0.79%) at 500 Hz and 6.44 Hz (0.64%) at 1 kHz. These values are slightly higher than those reported by Grant (1987) and Demany and Semal (1989) at the same frequencies, and consistent with FMDTs obtained at other frequencies by Strelcyk and Dau (2009), whose experimental procedure was used in the present study. In line with these three studies, there was a significant increase in absolute FMDT (or decrease if expressed as a percentage) with frequency in both groups of listeners (NH: p=0.031, HI: $p<0.001$, Wilcoxon paired signed-rank test).

With the exception of HI subjects 1, 7, 12, and 13, HI listeners showed elevated FMDTs, suggesting a deficit in monaural TFS processing. The group difference between NH and HI listeners was significant at both frequencies (500 Hz: p=0.005, 1 kHz: p=0.003), confirming the adverse effects of sensorineural hearing loss on low-rate FM detection found in earlier studies (*e.g.*, Lacher-Fougère and Demany, 1998; Moore and Skrodzka, 2002). Subjects 10 and 14 were among the HI listeners with the highest FMDTs at both test frequencies, and so were subjects 9 and 11, indicating that their deficit in TFS processing does not only reflect a specific binaural impairment. Moreover, mean FMDTs at 500 Hz were significantly correlated with MP$-$BP scores (p=0.014, ρ=0.69), further suggesting that impaired binaural-pitch perception mainly stems from a poor peripheral representation of fine temporal information. The correlation between FMDTs at 1 kHz and MP$-$BP scores was not significant (p=0.130), probably reflecting the fact that note frequencies in the pitch contour identification experiment did not exceed 800 Hz. Despite the above correlation, a clear-cut relationship between binaural-pitch perception and accurate monaural TFS processing cannot be established, as some listeners with highly elevated FMDTs did not show impaired binaural-pitch perception (subjects 2 and 5 at 500 Hz, subjects 4, 5, and 8 at 1 kHz). This might suggest that impaired monaural TFS processing is necessary, but not sufficient, to make binaural-pitch perception disappear. However, caution is needed in drawing such a conclusion, as it relies on the assumption that FM-detection at low rates exclusively relies on TFS processing. More central deficits unrelated to TFS processing may also play a role in the elevated FMDTs observed in some listeners (Lacher-Fougère and Demany, 1998), which is also suggested by the significant correlation of FMDTs at 500 Hz with MP+BP scores[3] (p=0.002, ρ=-0.75).

The asymmetry in the accuracy of peripheral TFS representation between ears is

---

[3] This correlation may at first suggest a role of TFS in pitch coding of MP stimuli as well as BP stimuli. However, one must keep in mind that the FM-detection task consisted in detecting the presence of a warble, which can be argued to involve a pitch cue. In that sense, it is possible that the FMDTs of some listeners, as well as the correlation of FMDTs with pitch contour identification scores, reflect a general ability of the listeners to follow pitch changes. Good performance in the FM-detection task might thus rely on more central pitch mechanisms, in addition to requiring accurate TFS processing.

another factor which may affect binaural-pitch perception. The relative asymmetry of individual FMDTs, expressed as the difference in thresholds divided by the mean left and right FMDT, is given in Table 3.1 (column “FM-As.”). Subject 10 obtained a particularly high asymmetry in FMDT. However, other listeners with highly asymmetric FMDTs were able to perceive binaural pitch (*cf.* subjects 2 and 8). Moreover, MP$-$BP scores were not correlated with FMDT asymmetry (500 Hz: p=0.194, 1 kHz: p=0.470) and there was no group difference in FMDT asymmetry between NH and HI listeners (p=0.627), suggesting that such asymmetries only have a limited or no effect on binaural pitch perception.

A significant correlation was found between mean FMDTs and BMLDs at 1 kHz (p=0.035, ρ=-0.57), but not at 500 Hz (p=0.192). However, the correlations of FMDTs with N_0S_π masked thresholds were significant at both frequencies (500 Hz: p=0.038, ρ=0.56; 1 kHz: p=0.021, ρ=0.61). Consistent with this, FMDTs showed correlations with IPD detection thresholds which were significant at 1 kHz (p=0.006, ρ=-0.69) and only borderline at 500 Hz (p=0.058, ρ=-0.52). This follows the observations of Strelcyk and Dau (2009), who found correlations between binaural and monaural measures of TFS processing in HI listeners. The correlation between FMDTs at 1 kHz and mean BILDs was also significant (p=0.018, ρ=-0.62).

Overall, binaural measures involving the use of interaural phase or time cues, both in noise and in quiet, may thus mainly rely on monaural TFS-processing skills. FMDTs at both frequencies were also correlated with reference SRTs (500 Hz: p=0.035, ρ=0.57; 1 kHz: p=0.046, ρ=0.54). This is in line with the results of Buss *et al.* (2004), who found a similar correlation, suggesting a role of TFS processing deficits in the reduced speech reception of listeners with sensorineural hearing loss. Finally, FMDTs were significantly correlated with hearing thresholds (500 Hz: p=0.012, ρ=0.47; 1 kHz: $p<0.001$, ρ=0.67). This reflects the heterogeneity of the listeners in terms of their audiograms, and is in line with the significant correlation of FMDTs with hearing thresholds reported by Lacher-Fougère and Demany (1998). This does nevertheless not exclude the presence of suprathreshold deficits, as listeners with similar audiograms may exhibit widely different FMDTs (Strelcyk and Dau, 2009).

3.9 Frequency selectivity

A notched-noise paradigm (Patterson and Nimmo-Smith, 1980; Patterson and Moore, 1986) was used to derive auditory filter shapes at 500 Hz in the subjects' left and right ears. Only six of the NH subjects participated in this experiment. The best-fitting rounded-exponential filter was estimated using the roex(p_u, p_l, r) filter model (Patterson *et al.*, 1982; Glasberg and Moore, 1990). The average rms fitting error over all ears was 0.67±0.24 dB, indicating reasonable fits provided by the model. The estimated auditory-filter equivalent rectangular bandwidths (ERB, Glasberg and Moore, 1990) for individual ears are given in the right panel of Fig. 3.7. The mean auditory filter bandwidth in the NH group was 85.9 Hz. Considering the low number of subjects and the resulting standard deviation (12.3 Hz), this is in good agreement with the value of 78.7 Hz suggested by the formula of Glasberg and Moore (1990), as well as the results of Moore *et al.* (1990) who reported a mean value of 87 Hz at 400 Hz. The ERB values were significantly higher in the HI than in the NH group ($p<0.001$), by a factor of 1.4 on average. Subjects 10 and 14 both showed very broad filters. However, MP$-$BP scores were not correlated with mean ERB values ($p=0.397$). This supports earlier suggestions that reduced frequency selectivity coexists with impaired binaural pitch perception, but cannot alone account for it (Santurette and Dau, 2007; Nitschmann *et al.*, 2010). Mean ERB values were not correlated with MP+BP scores ($p=0.923$).

The relative asymmetry between left and right ERB values, expressed as the difference in ERB divided by the mean left and right ERB, is given in Table 3.1 (column "ERB-As."). Most HI listeners did not show higher ERB asymmetry than NH listeners. Only subject 10 had an asymmetry factor more than double that of the NH group. This may be a contributing factor to her reduced performance in binaural TFS processing measures, and hence her inability to perceive binaural pitch. The outputs of left and right filters with different bandwidths may indeed show reduced correlation, leading to less effective binaural unmasking. However, subject 14 showed little asymmetry, and the lack of correlation between ERB asymmetry and MP$-$BP scores ($p=0.564$), BMLDs ($p=0.062$), and mean BILDs ($p=0.731$), suggests that ERB asymmetry was not a crucial contributor to the observed deficits in binaural TFS processing.

The auditory filter bandwidths were significantly correlated with hearing thresholds at 500 Hz (p=0.029), confirming the relationship between audibility and frequency resolution found in earlier studies (*e.g.*, Tyler *et al.*, 1983; Strelcyk and Dau, 2009). However, the mean ERB values did neither correlate with speech reception (p=0.329), mean BILDs (p=0.239), IPD detection thresholds (p=0.066), nor FMDTs (500 Hz: p=0.542, 1 kHz: p=0.128). The only other measure with which mean ERB values showed a significant correlation was the BMLD (500 Hz: p=0.017, ρ=-0.64; 1 kHz: p=0.005, ρ=-0.72). Moreover, the examples of subjects 5 and 6, who showed relatively narrow filters but elevated FMDTs, and of subjects 1, 7, and 13, who obtained normal FMDTs despite broader filters, illustrate that deficits in frequency selectivity and TFS processing do not necessarily covary. Therefore, the present findings are in line with those of Strelcyk and Dau (2009), providing further evidence that deficits in TFS processing cannot entirely be accounted for by poor frequency selectivity and reflect an additional impairment specific to the temporal acuity of internal sound representations.

3.10 Overall summary and discussion

3.10.1 Binaural pitch perception and the auditory profile

Eight NH listeners and fourteen HI listeners with sensorineural hearing loss and various audiometric configurations performed a pitch contour identification task with binaural pitch stimuli and salience-matched monaurally-detectable pitches. While most HI listeners could detect both pitch types as often as NH listeners, two of them were found not to perceive binaural pitch at all. Pitch contour identification scores showed that binaural pitch was clearly audible for all other listeners, but significantly less salient than the monaurally-detectable pitch. This indicates that the impaired mechanisms in sensorineural hearing loss affect pitch perception of noise-based pitch-evoking stimuli to a larger extent if pitch extraction requires binaural processing. The controlled pitch detection and contour identification task used here was found to be a more reliable test for the ability to hear binaural pitch than a short scale test (Santurette and Dau, 2007), in which a negative response may not imply absent binaural pitch perception.

In an attempt to better characterize the hearing loss of the individual listeners, tympanograms, acoustic reflexes, and otoacoustic emissions were measured, and loudness perception was evaluated using a categorical loudness scaling procedure and a SISI test. In one of the listeners without binaural pitch perception, the results from these tests were consistent with a retrocochlear disorder, while the second listener who could not hear binaural pitch showed signs of a cochlear impairment, possibly combined with further retrocochlear damage. Among the remaining HI listeners, who all clearly perceived binaural pitch, some showed strong signs of cochlear impairments, while others had no evident abnormal cochlear function. Overall, the results from these tests, together with the unavailability of clinical diagnoses for the individual listeners, did not allow clear conclusions about the sites of the impairment, such that no link could be established between the absence of binaural pitch percept and the presence of a retrocochlear disorder.

The outcome of the binaural pitch experiments was compared to the listeners' performance in measures of several specific auditory and cognitive functions. The 'BP' entry in Table 3.3 summarizes the relationship between binaural pitch perception and these auditory-profile measures[4]. The results from a lexical decision task and a reading span test showed that cognitive abilities were unrelated to binaural pitch perception. However, deficits in binaural pitch perception coexisted with a large reduction in binaural release from masking, according to BMLD and BILD estimations. This nearly absent binaural advantage in detecting tones or understanding speech in background noise was linked to a more fundamental deficit in the processing of interaural phase information in quiet, illustrated by very low upper-frequency limits

---

[4] For the 'BP' entry in Table 3.3, '+' indicates a significant correlation of MP−BP scores with the corresponding measure at 500 Hz, while '—' indicates the absence of any correlation between MP−BP scores and the corresponding measure. As MP−BP scores were not calculated for the two listeners who could not detect binaural pitch, '+b' is used to indicate that these subjects, as well as subjects with low MP−BP scores, showed considerably reduced performance in the corresponding tests.

for detecting a large IPD. Such IPD detection thresholds were clearly related to the ability of the listeners to hear binaural pitch in the frequency range where it is most salient. This suggests that accurate TFS processing up to the stage of binaural integration is a crucial factor for a binaural pitch sensation to arise. Although the listeners with absent binaural pitch perception exhibited broad auditory filters, reduced frequency selectivity did not account for the reduced salience of binaural pitch in HI listeners, suggesting a primary role of TFS processing.

At the output of the binaural processing stage, the accurate representation of precise timing information relies on two factors: a good peripheral temporal acuity as conveyed *via* phase-locking, and an accurate comparison of temporal inputs from the left and right channels *via* a well-functioning binaural processor. As both factors can affect performance in the IPD detection task, an additional monaural measure thought to primarily rely on peripheral phase-locked information, FM detection at a low FM-rate, was carried out. Here again, the listeners with absent binaural pitch perception performed markedly more poorly than other HI listeners, suggesting impaired peripheral representations of TFS in listeners unable to hear binaural pitch. A similar degree of monaural TFS-processing deficit was, however, also found in some listeners with immediate binaural pitch perception, suggesting that the reduced acuity of phase-locked temporal information in the periphery is not a sufficient factor to make binaural pitch perception break down completely. Despite this, the significant correlation observed between 500-Hz FMDTs and the difference in pitch identification scores for MP *vs.* BP stimuli supports the existence of a link between the acuity of peripheral TFS processing and the salience of binaural pitch.

In summary, binaural pitch perception was clearly found to primarily rely on TFS processing abilities, but it remains uncertain whether the peripheral or central mechanisms involved in processing fine temporal information are most crucial for binaural pitch extraction. Only a study on a large number of subjects with specific diagnoses might further reveal whether the absence of binaural pitch percept is a valid indicator of a particular auditory disorder.

Table 3.3: Summary of the relationship between outcomes of the different experiments. '++': Significant correlation. '+' Significant correlation for at least one test frequency. '+b': Reduced performance in listeners with impaired binaural pitch perception. '—': No relationship found. "BP" refers to binaural pitch perception[3], "Cog." to cognitive abilities, and "Aud." to audibility.

	BP	Cog.	SRT	ILD	BILD	BMLD	IPD	FM	ERB	Aud.
BP		—	—	+b	+b	+b	+b	+	—	—
Cog.	—		—	—	—	—	—	—	—	—
SRT	—	—		++	—	++	—	++	—	++
ILD	+b	—	++		—	—	—	—	—	++
BILD	+b	—	—	—		+	—	+	—	—
BMLD	+b	—	++	—	+		++	+	++	+
IPD	+b	—	—	—	—	++		+	—	++
FM	+	—	++	—	+	+	+		—	++
ERB	—	—	—	—	—	++	—	—		++
Aud.	—	—	++	++	—	+	++	++	++	

3.10.2 Correlations between the auditory profile measures

Overall, the listeners from the HI group showed significant deficits in masked detection of homophasic and antiphasic tones, speech intelligibility in background noise, spatial release from masking, binaural and monaural TFS processing, and frequency selectivity. While a subgroup of HI listeners showed reduced binaural advantage in noise, others could benefit from binaural processing to the same degree as NH listeners for detecting dichotic tones in noise or understanding speech in lateralized noise. Measures of lexical decision and working memory processing and capacity did not reveal highly significant differences between NH and HI listeners. Furthermore, cognitive abilities were not correlated with any of the measures included in the auditory profile. The only listener suffering from obscure dysfunction showed reduced speech reception in noise compared to NH listeners. However, no clear deficit was found for this listener in any of the other auditory-profile tests.

The study of correlations between the different auditory-profile measures, summarized in Table 3.3, revealed an important role of monaural TFS processing ('FM' entry) for a variety of other tasks, including the ability to process interaural phase disparities both in quiet and in noise, but also to identify pitch contours and understand speech in background noise. The correlations between monaural TFS processing abilities and the performance in the binaural tasks do not support the presence of a specific binaural component in sensorineural hearing loss. This conclusion nevertheless relies on the assumption that elevated FMDTs exclusively reflect a monaural TFS deficit, which remains controversial (Lacher-Fougère and Demany, 1998). Moreover, the listeners who could detect a 180ř IPD might still have difficulties in detecting smaller IPDs. Consequently, good performance in the IPD detection experiment does not rule out the presence of additional deficits in binaural TFS processing. Asymmetries between ears in the different monaural measures were generally not found sufficient to account for reduced performance in binaural tasks, even though they might be a contributing factor.

Most notable was the absence of correlation between auditory filter bandwidths and TFS-related measures, strongly suggesting that the TFS-processing deficits observed in HI listeners are at least partly independent of frequency selectivity. While the effect of a loss of OHCs on phase-locking precision remains uncertain, the loss of IHCs and auditory nerve fibers seem the most plausible factors likely to affect the peripheral representation of TFS information (Moore, 2007; Strelcyk and Dau, 2009). However, the extent to which damage in each of these entities contributes to impaired TFS processing remains difficult to quantify.

The heterogeneity of the audiogram shapes in the HI group should be mentioned as a possible confounder in the present study, as it resulted in the correlation of most measures with audibility ('Aud.' entry in Table 3.3). Particularly listeners with elevated low-frequency hearing thresholds often showed markedly poorer performance than most other HI listeners, leading to significant correlations with low-frequency audibility. Although the use of loudness-adjusted stimuli aimed at limiting the influence of low sensation levels on the different outcomes, such an influence cannot be excluded here, as binaural performance is increasingly affected by presentation level as the latter approaches hearing threshold (*e.g.*, Hershkowitz and Durlach, 1969). However, the presence of recent counterexamples indicates that audibility cannot alone account for a loss of binaural pitch percept (Nitschmann *et al.*, 2010), and there is evidence of suprathreshold deficits in HI listeners, at least concerning TFS processing, from studies on homogeneous groups of listeners in terms of audibility (*e.g.*, Strelcyk and Dau, 2009).

No clear distinction could be made between listeners with supposed cochlear *vs.* retrocochlear hearing losses in any of the auditory-profile tests. While the lack of formal diagnoses and the difficulty of interpreting audiometric and loudness measures prevented a clear classification of the listeners in such subgroups, one should also keep in mind that each auditory-profile measure might be prone to several factors from different sites of impairment. The relative contribution of hair-cell *vs.* nerve-cell loss on spectral and temporal resolution is an obvious example of this. If the need for reliable site-of-lesion tests persists, the present findings may help to define which basic features of hearing are primarily at stake and how they relate in cases of sensorineural hearing loss. Specifically, they underline the presence of TFS processing deficits which cannot be fully accounted for by a loss of frequency selectivity, and may adversely affect speech and pitch perception in background noise. Consequently, the evaluation of TFS processing abilities in HI patients would seem a valuable addition to audiometric measures and an informative tool in terms of general hearing abilities. In contrast, a measure of frequency selectivity may be redundant in a time-constrained context, due to the observed correlations with audibility, and may reflect the difficulties of HI listeners in other tasks to a lesser extent.

The present findings are thus in line with the increasing evidence for an important and independent role of TFS processing in hearing (*e.g.,* Lacher-Fougère and Demany, 2005; Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008; Strelcyk and Dau, 2009). However, this remains a controversial issue. One reason for this is the current lack of a reliable non-invasive measure of peripheral TFS processing in humans. Uncertainties persist concerning the role of a temporal mechanism for FM detection at very low rates (Lacher-Fougère and Demany, 1998), the measure chosen in the present study. Another disputed method is the discrimination of harmonic and frequency-shifted bandpass-filtered complex tones (Moore *et al.*, 2009a), for which the sole role of TFS information has been questioned (Oxenham *et al.*, 2009), and which cannot be used at low frequencies. The search for a psychophysical or physiological outcome that would accurately reflect peripheral TFS-processing abilities thus ought to be pursued. Until this is achieved, given the observed dependence of binaural TFS-processing outcomes on peripheral phase-locking acuity, the binaural IPD detection test used here may be a useful tool to quickly reveal deficits in the use of low-frequency TFS cues. The latter test proved a fast and reliable[5] measure with a relatively simple task. This is unlike most auditory-profile tests used in the present study, which were well suited to a laboratory study on a few listeners but, with the exception of audiometric and loudness-perception measures, would have been too time-consuming in a clinical set-up. Therefore, further efforts to design an adequate battery of tests would be beneficial. Ideally, such a battery should allow an evaluation of each subject's auditory profile, in detail, but quickly, and without redundancy. Whether the inclusion of a binaural-pitch test would be a valuable addition to another measure of binaural TFS processing such as IPD detection could not be determined here. In this respect, the outcome of a short binaural-pitch test on a large population of listeners with confirmed specific sites of impairments would be very informative.

Appendix 3: Methods for the auditory profile tests

Pure-tone audiometry, tympanometry, acoustic reflex (sections 3.4.1 and 3.4.2): Air-conduction thresholds were obtained in both ears at each of the following frequencies: 125, 250, 500, 1000, 2000, 3000, 4000, and 8000 Hz. Bone-conduction thresholds were obtained in both ears at 500, 1000, 2000, and 4000 Hz. All audiograms were measured using the Interacoustics AC440 audiometry module for the Affinity hearing-aid analyzer. An Interacoustics AT235 impedance audiometer was used to evaluate middle-ear function and acoustic reflexes.

Otoacoustic emissions (section 3.4.3): The click stimuli were generated in MATLAB, sent to an RME FireFace 800 A/D-D/A converter *via* the pa-wavplay software, and presented to the test subjects *via* an ER-2 probe at a rate of 20 clicks per second. The signal level of 70 dB peSPL was controlled with a DT-PA5 programmable attenuator. Recordings were made using an ER-10B low-noise microphone, and were

---

[5] For the IPD-detection test, the within-subject standard deviation over all measurement blocks was on average 1% of the mean log threshold, indicating a good measurement repeatability. This is in contrast to the FM-detection test, for which the uncertainty around the individual thresholds is higher, with a median block standard deviation of 15% of the mean threshold.

bandpass filtered between 0.6 and 5 kHz using an analog Rockland 852 HI/LO filter. The recorded analog signals were then converted and stored digitally, and the final click-response was defined as the average of 2000 recordings. Test subjects were instructed to lie down in a soundproof booth and keep still.

Loudness scaling (section 3.5.1): The stimuli were one-third octave bands of low-noise noise, geometrically-centered at the test frequency. They were generated using Method 1 as in Kohlrausch *et al.* (1997), and a stimulus duration of 1 second was used. The categorical scale contained 11 response alternatives (ISO 16832, 2006) and the subjects were instructed to click on one of the response bars after each presentation. Other test subject instructions, as well as the adaptive procedure used, were the same as in Brand and Hohmann (2002), except that the starting level and maximum level were 65 dB SPL and 115 dB SPL respectively, and that three iterations were used in the second phase, with estimated levels L_{10}, L_{20}, L_{30}, and L_{40} presented in the third iteration. After each block, the loudness function was obtained by fitting two independent linear functions to the lower ($L \leq L_{15}$) and upper ($L \geq L_{35}$) sections of the data. A Bézier smoothing was then used to link the two linear ends of the curve (Brand and Hohmann, 2002). When such a fit was not possible, the lower and upper sections of the curve were extended to $L \leq L_{20}$ and $L \geq L_{30}$, respectively, before the Bézier smoothing was applied. Further extensions of the linearly-fitted lower and upper ends of the curve, in steps of 5 cu, were used in some cases. If the variation in the subject's ratings was too large to obtain a loudness curve with this fitting method, the measurement was repeated.

SISI test (section 3.5.2): The automatic procedure of the Interacoustics AC222 audiometer was used. A continuous pure tone was presented at a level of 20 dB SL, and short intensity increments of 5, 2, or 1 dB occured at periodic time intervals. The task of the listeners was to press a response button every time an intensity increment was heard. The SISI score was calculated as the percentage of 1-dB increments correctly detected by the subject. Twenty 1-dB increments were presented for each condition.

Cognitive tests (section 3.6): Danish versions of both tests were implemented using the Psychophysics Toolbox extensions for MATLAB (Brainard, 1997; Pelli, 1997). Subjects were seated in front of a computer screen, such that the centre of the screen was at eye level and the distance from their eyes to the screen was approximately 50 cm. In both tests, words were presented in white bold capital letters on a dark green background at the centre of the screen. Subjects were instructed to keep their attention to the centre of the screen at all times, and a fixation point in the shape of a square appeared before each presentation. All room lights were switched off during testing.

Lexical decision (section 3.6.1): Four lists containing 50 words each (25 real words and 25 non-words) were created, as well as a 10-word practice list. All words contained 3 to 5 letters, and both started and ended with a consonant. Real words were monosyllabic adjectives or nouns, all situated between the 1000th and 2000th most frequent words in Danish (1998-2002), as selected from the Korpus 2000 database (DSL, 2002). Non-words were non-homophonic, phonologically different but visually similar to Danish language (pseudowords), and were obtained by modifying one letter from real Danish words. Each test list contained 9 three-letter, 12 four-letter, and 4 five-letter real words and non-words. Subjects gave their responses *via* a computer keyboard, and were instructed to place their right index on the 'K' key, and their left index on the 'F' key before the experiment started. Their task was to press 'K' (korrekt/correct) as soon as a real word appeared on the screen and 'F' (forkert/wrong) as soon as a non-word appeared. Before each word presentation, the fixation point was shown for a randomly chosen period of minimum 2 and maximum 4 seconds. No feedback was provided. Each subject performed a single block on one test list, after one training session with the practice list. Words in each list were presented in a random order.

Reading span (section 3.6.2): One list containing 54 four-word sentences (27 normal and 27 absurd) was created, as well as a 6-sentence practice list. Absurd sentences were all grammatically correct. Each sentence was presented word by word, and each word was visible for a fixed period of 800 ms. After the last word of a sentence was shown, a question mark appeared at the centre of the screen. In the same manner as in the lexical decision task, subjects were instructed to press ‘K’ (korrekt/correct) if the sentence was normal and ‘F’ (forkert/wrong) if the sentence was absurd, as soon as the question mark appeared. Before each sentence presentation, the fixation point was shown for a randomly chosen period of minimum 2 and maximum 4 seconds. No feedback was provided. After the presentation of a group of 3 to 6 sentences, the word “FØRST” (first) or “SIDST” (last) appeared on the screen, and the subjects were asked to verbally recall either the first or the last word of each sentence in the group. A voice-recording device was used to gather the subjects’ responses. Subjects were allowed to guess and report words in any order, and had unlimited time to answer. The number of sentences in each group increased during the test, starting with three groups of 3 sentences, followed by three groups of 4, 5, and finally 6 sentences. First words had to be recalled for half of the groups, last words for the other half, and the sequence of “first” and “last” groups was randomized. Each subject performed a single test block, after one training session with the practice list. Sentences were presented in a random order.

BMLD (section 3.7.1): In all conditions, the 2-octave wide noise was geometrically centered around the test frequency. The noise level was fixed and equal to the binaurally-measured MCL at the test frequency. The signal level was varied adaptively in a 3-interval, 3-alternative forced-choice (3I-3AFC) paradigm. Intervals had a 500-ms duration, including 50-ms onset and offset cosine-ramps, and were separated by a 333-ms silent pause. A 1-up 2-down procedure was used, tracking the 70.7% point on the psychometric function (Levitt, 1971). For each presentation, one randomly chosen interval contained the tone signal in background noise, while the two other intervals contained noise only. The task of the subjects was to indicate *via* a computer keyboard which interval contained the tone. Feedback was provided. The starting signal-to-noise ratio (SNR) was 0 dB. Stepsizes of 8, 3, and 1 dB were used, and the stepsize was decreased after each upper reversal. A block was terminated after 10 reversals and the threshold value was determined from all points following the 4$^{\text{th}}$ reversal. Each subject performed three blocks with each condition, including one practice block. The final threshold was defined as the average threshold over the last two blocks.

BILD (section 3.7.2): The interfering noise consisted of superimposed speech material from the DANTALE II sentences, yielding optimal spectral masking (Wagener *et al.*, 2003). The spatial origin of the speech and noise stimuli was controlled virtually by convolving the waveforms with a set of HRTFs, as measured by Gardner and Martin (1994). Each subject was first introduced to the procedure and the speech material in a training block consisting of 30 sentences in the reference condition. A single test block of 20 sentences was then performed for each condition. In each block, the noise level was kept constant at the binaurally-measured MCL at 500 Hz, and the speech level was varied adaptively. Sentence lists were chosen randomly and the subjects' task was to verbally report the words in each sentence as they were understood, after each presentation. For each condition, the SRT was defined as the SNR for which 50% of individual words were correctly identified.

IPD detection (section 3.8.1): A 3I-3AFC procedure was used, in a paradigm similar to that of Ross *et al.* (2007b). Stimuli were sinusoidal-amplitude-modulated pure tones, with a 40-Hz modulation rate and a modulation depth equal to 1. The tracking variable was the frequency of the tone carrier. For each trial, three 750-ms intervals separated by 333-ms silent gaps were presented. In the two reference intervals, the left and right stimuli were in phase, and were perceived as a single sound source located inside the head. In the randomly chosen target interval, the left and right stimuli were in phase during the first half (375 ms), and in antiphase during the second half of the interval, *i.e.*, the sound was perceived as starting inside the head and suddenly becoming more spacious in the middle of the interval. As the modulation rate used corresponded to a 25-ms envelope period, the change in IPD always occured in a modulation dip, thus avoiding discontinuities in the waveform (Ross *et al.*, 2007b). The task of the subjects was to indicate *via* a computer keyboard which interval contained the IPD change. Feedback was provided. A 2-up 1-down procedure was used to track the 70.7% point on the psychometric function (Levitt, 1971). The threshold frequency was tracked logarithmically, with an initial carrier frequency of 250 Hz and stepsizes of 1/2, 1/5, and 1/10 octave, which were decreased after each lower reversal. If a subject could not detect the target interval correctly in the first trial, the initial carrier frequency was reduced to 100 Hz. A block was terminated after 10 reversals and the threshold value was determined from all points following the 4th reversal. Each subject performed five test blocks and the final threshold was defined as the average threshold over all blocks. A presentation level of 50 dB SPL was used in NH subjects. In order to ensure sufficient audibility of the stimuli for HI listeners, the presentation level was adjusted to the higher of the binaurally-measured L_{10} levels at 500 Hz and 1000 Hz (Table 3.1), when these were higher than 50 dB SPL. If the latter adjustment was insufficient to reach a level of 20 dB SL in both ears, the level corresponding to 20 dB SL in the worst ear was used. Although asymmetries in hearing threshold might disrupt the perceived location of the stimulus by introducing interaural level differences (ILDs), results from pilot testing showed that the IPD cue was still easily perceivable when an ILD of 15 dB or less was introduced. Moreover, an IPD of 180° is much larger than the just noticeable difference in interaural phase for a 500-Hz tone in NH listeners, for ILDs up to at least 20 dB (Hershkowitz and Durlach, 1969; Domnitz, 1973). Therefore, such an IPD is expected to be easily detectable in the presence of an ILD. Hearing threshold asymmetries were thus only compensated for by introducing an ILD in the stimulus when the mean asymmetry between 125 and 2000 Hz was equal to or larger than 15 dB (Table 3.1).

FM detection (section 3.8.2): The stimuli and procedure were similar to those used by Strelcyk and Dau (2009). An FM-rate of 2 Hz and an FM-phase of 1.5π were used, such that the frequency-modulated stimulus could be described as

$$x(t) = a(t)\sin\left[2\pi f_c t + \frac{\Delta f}{2}\sin(\pi(4t+1.5))\right],$$

where f_c is the carrier frequency and Δf the maximum frequency excursion. Quasi-sinusoidal amplitude modulation (AM) was superimposed to the FM-tones in order to disrupt FM-to-AM conversion cues (Grant, 1987; Moore and Sęk, 1996), such that $a(t)$ was proportional to $1 + m\sin(2\pi F(t) + \Phi)$. The AM-depth m was fixed at a peak-to-valley ratio of 6 dB and the AM-phase Φ was randomized. The integral of the instantaneous AM-rate was

$$F(t) = \int_0^t d\tau \left(f_1 + \frac{f_2 - f_1}{T}\tau\right),$$

where T is the stimulus duration and f_1 and f_2 were randomly chosen between 1 and 3 Hz with $|f_2 - f_1| > 1$ Hz. An adaptive 3I-3AFC procedure was used, with an interval duration of 750 ms, and 300 ms silent gaps between two successive intervals. All stimuli were gated with 50 ms $\cos^2$-ramps. In each trial, all three intervals were independently amplitude modulated while only the target interval was frequency modulated. The subjects' task was to indicate *via* a computer keyboard which interval contained the FM-tone. The 75% point on the psychometric function was tracked using a weighted up-down method (Kaernbach, 1991) in which Δf was varied logarithmically. A block was terminated after 12 reversals and the threshold value was determined from all points following the 4th reversal. Each subject participated in a training session containing 2 blocks in each ear/test-frequency condition. Three test blocks were then performed for each condition. Additional blocks were performed as long as the standard deviation over all blocks exceeded 15% of the mean FMDT, with a maximum of 5 blocks per subject. The final threshold was defined as the geometric mean over all blocks. The presentation order of the 4 different ear/test-frequency conditions was randomized for each subject. A 60 dB SPL presentation level was used, unless the subject's left or right L_{10} level at 500 Hz or 1000 Hz was higher than 60 dB SPL, in which case the higher of the left and right monaurally-measured L_{10} at 500 and 1000 Hz was used (Table 3.1). If the latter adjustment was insufficient to reach a level of 20 dB SL in both ears, the level corresponding to 20 dB SL in the worst ear was used.

Frequency selectivity (section 3.9): Target tones of 440-ms duration were temporally centered in 550 ms fixed-amplitude random-phase noise maskers, and 50 ms $\cos^2$-ramps were applied to both tones and maskers. The outside edges of the noise maskers were fixed at $\pm 0.8 f_0$, where f_0 is the signal frequency. Five symmetric ($\Delta f / f_0 = [0.0|0.0; 0.1|0.1; 0.2|0.2; 0.3|0.3; 0.4|0.4]$) and two asymmetric ($\Delta f / f_0 = [0.2|0.4; 0.4|0.2]$) notch conditions were measured, where Δf is the spacing between f_0 and the inner noise edges. An adaptive 3I-3AFC weighted up-down method (Kaernbach, 1991) was used, tracking the 75% point on the psychometric function. Successive intervals were separated by a 250 ms silent gap. In each trial, all three intervals contained the noise masker, while only the target interval contained the tone signal. The subjects' task was to indicate *via* a computer keyboard which interval contained the tone. The masker level varied adaptively while the signal level was kept constant at 50 dB SPL, unless the subject's left or right L_{10} level at 500 Hz was higher than 50 dB SPL, in which case the monaurally-measured L_{10} in the worst ear was used (Table 3.1). A block was terminated after 12 reversals and the threshold value was determined from all points following the 4th reversal. Each subject performed three test blocks in each condition, and the final threshold was defined as the average threshold over all blocks. The presentation order of the 14 different ear/notch conditions was randomized for each subject.

Acknowledgments

Special thanks to Sarah Verhulst, for her help with setting up and analysing OAE measurements, Jens Bo Nielsen, for proofreading the Danish words and sentences used in the cognitive tests, and Olaf Strelcyk, for providing the code used in the FM-detection and notched-noise experiments.

4

Detection and identification of monaural and binaural pitch contours in dyslexic listeners[‡]

The use of binaural pitch stimuli to test for the presence of binaural auditory impairment in reading-disabled subjects has so far led to contradictory outcomes. While some studies found that a majority of dyslexic subjects was unable to perceive binaural pitch, others obtained a clear response of dyslexic listeners to Huggins' pitch (HP). The present study clarified whether impaired binaural pitch perception is found in dyslexia. Results from a pitch contour identification test, performed in 31 dyslexic listeners and 31 matched controls, clearly showed that dyslexics perceived HP as well as the controls. Both groups also showed comparable results with a similar-sounding, but monaurally detectable, pitch-evoking stimulus. However, nine of the dyslexic subjects were found to have difficulty identifying pitch contours both in the binaural and the monaural conditions. The ability of subjects to correctly identify pitch contours was found to be significantly correlated to measures of frequency discrimination. This correlation may be attributed to the similarity of the experimental tasks and probably reflects impaired cognitive mechanisms related to auditory memory or auditory attention rather than impaired low-level auditory processing *per se*.

[‡] This chapter is based on Santurette *et al.* (2010).

4.1 Introduction

Developmental dyslexia is a specific learning impairment affecting the ability to fluently read, spell, and decode words, despite adequate educational opportunities and otherwise normal intellectual abilities (Lyon *et al.*, 2003). The basis for this disorder, estimated to affect 5 to 10% of school-aged children (Shaywitz *et al.*, 1990), is generally described by theories supporting two antagonistic points of view (see *e.g.*, Ramus (2003); Rosen (2003) for reviews). In short, the discussion revolves around whether the phonological processing disorder found in dyslexia is due to a purely cognitive deficit, linked to congenital dysfunction in the corresponding cortical areas of the brain, or to a basic sensorimotor deficit, possibly linked to a dysfunction of magno-cells along the sensory pathways. Conclusions favoring both points of view have been drawn from empirical data, and a possible relationship between low-level nonlinguistic impairment and reading disability remains under debate.

Although their influence on reading abilities is controversial, a wide range of auditory processing disorders have been found in part of the dyslexic population (*e.g.* Wright *et al.*, 2000; Amitay *et al.*, 2002). Among other things, it has been suggested that low-level binaural processing might be impaired in some dyslexic listeners. McAnally and Stein (1996) obtained significantly lower binaural masking level differences (BMLDs) in their group of dyslexic listeners than in the control group, suggesting a difficulty of dyslexic listeners in exploiting interaural phase differences to obtain a binaural advantage. However, later studies (Hill *et al.*, 1999; Amitay *et al.*, 2002) found similar BMLDs in dyslexics and controls. Binaural pitch stimuli (Cramer and Huggins, 1958) have also been used to investigate binaural processing abilities in dyslexic listeners. Dougherty *et al.* (1998) found that most of their dyslexic subjects failed at identifying and lateralizing pitch contours if no monaural cues were available. Their results suggested that the binaural integration of fine temporal information might be impaired in dyslexia, thus inducing an inability to perceive binaural pitch. Similarly, Edwards *et al.* (2004) asked a group of reading-disabled children to lateralize binaural pitch stimuli, and found that 52% of dyslexics failed at the task in the absence of monaural cues. In contrast, Chait *et al.* (2007) did not find evidence for impaired binaural pitch perception in dyslexia. In a pitch onset detection task, they compared the detectability of Huggins' pitch (HP) to that of sinusoidal targets in diotic noise (TN) and found that HP was generally perceivable by dyslexic listeners, with only few misses on average. Moreover, elevated response times with both HP and TN stimuli suggested a slower processing of pitch-evoking noise stimuli in dyslexics, rather than an impairment in pitch detectability *per se*. Short stimulus durations or high task complexity might then have been responsible for the results obtained by Dougherty *et al.* (1998) and Edwards *et al.* (2004).

The question remains whether some reading-disabled listeners have impaired binaural pitch perception, or whether all of them are able to hear binaural pitch, provided the duration of the stimuli is sufficiently long and the task simple enough. The present study aimed at clarifying this point, by investigating the ability of a larger group of dyslexic listeners to detect and identify binaurally- and monaurally-detectable pitch contours, using two different stimulus durations. By comparing the subjects' detection scores to their pitch contour identification scores, and by evaluating correlations of the results with specific auditory and cognitive measures, the study aimed to verify or falsify the presence of low-level binaural processing impairment in dyslexia.

4.2 Methods

4.2.1 Procedure

A pitch contour identification test was performed with two stimulus types eliciting a pitch sensation in noise: a binaural pitch (BP) stimulus, and a similarly-sounding stimulus containing a monaurally-detectable pitch (MP) (see section 4.2.3). The use of two stimulus types was motivated by two factors: (a) Assuming that all listeners could perceive MP, it made pitch contour identification measureable in subjects unable to perceive BP; (b) It allowed evaluation of whether potential difficulties in pitch detection and contour identification were linked to a deficit in binaural processing or to a general difficulty in extracting tonal objects from background noise (*cf.* Chait *et al.*, 2007). The fundamental difference between these two stimuli is that, while BP requires binaural presentation and cannot be perceived when listening with only one ear (in which case only noise is heard), MP can be detected monaurally.

In each trial, sequences of three musical notes were presented, such that they formed either rising, falling, or constant pitch contours (Tables 4.1(a),4.1(b)). Note frequencies were chosen to be between 500 and 800 Hz, *i.e.*, within the range of strongest salience of Huggins' pitch (Santurette and Dau, 2007). The choice of rather large frequency intervals (>17%) between successive notes in the rising and falling pitch contours (Table 4.1(c)) was made to avoid possible effects of impaired frequency discrimination on pitch contour identification: France *et al.* (2002) obtained just noticeable differences (JNDs) that never exceeded 16% of the test frequency (500 Hz) in their study of auditory frequency discrimination in dyslexia, and normal frequency sensitivity usually lies around 1-3% in the frequency range considered here (Moore, 2003). Two different note durations were used in order to measure whether performance in pitch detection and contour identification improved with stimulus duration: As response times obtained by Chait *et al.* (2007) were in the range 400-800 ms for both dyslexics and controls, note durations of 300 ms (shorter than the subjects' response time) and 900 ms (which should be long enough for all subjects to extract the pitches from the noise) were compared.

Table 4.1: Note frequencies, pitch contours, and frequency intervals used in the pitch contour identification experiment. (a) Note frequencies. (b) Pitch contours. (c) Absolute and relative frequency intervals between successive notes. Relative intervals between f_1 and f_2 are given in percent relative to $(f_2+f_1)/2$.

(a) Note frequencies

Note	Frequency
C_5	523.25 Hz
E_5	659.26 Hz
G_5	783.99 Hz

(b) Pitch contours

Contour	Note sequence
Rising	C_5-E_5-G_5
Falling	G_5-E_5-C_5
Constant	E_5-E_5-E_5

(c) Frequency intervals

Notes	Interval
C_5-E_5	136 Hz (23%)
E_5-G_5	125 Hz (17%)

After each presentation, the subjects responded by pressing one of four buttons on a computer screen: an upward-pointing arrow (rising pitch), a downward-pointing arrow (falling pitch), a horizontal arrow (constant pitch), or a cross (no pitch). Subjects were instructed to press the cross when no melody was heard, and to press the arrow corresponding to the perceived pitch contour when a melody was heard. The "no melody" option was included so that both detection and contour recognition could be tested within a single short experiment. Subjects were presented 15 trials for each combination of stimulus type (MP or BP) and note duration (300 or 900 ms). In addition to these 60 trials containing a pitch contour, 20 trials containing no pitch contour (diotic white noise only) were presented, half of them corresponding to each duration. This made it possible to evaluate false alarms, and to avoid the possibility of subjects never pressing the cross. Trials were presented in a random order, and total testing time was *ca.* 10 minutes per subject. Before the test, each subject was first introduced to the different pitch contours played with pure-tone stimuli. A short 12-trial practice run was also performed with pure tones to ensure that the task was correctly understood.

4.2.2 Subjects

Two groups of 31 dyslexic subjects (ages: 19-30 years, mean: 21.5) and 31 matched controls (ages: 19-32 years, mean: 21.4) with normal hearing thresholds participated in the experiment. All experiments were approved by the Committee of Medical Ethics of Clinical Research of the Katholieke Universiteit Leuven. Subjects had Dutch as a native language and were matched according to gender, age, and educational level. The cognitive profiles of the two groups of subjects are summarized in Table 4.2(a). Dyslexics performed significantly worse than controls in all measures of reading and spelling accuracy, rapid automatized naming, phonemic awareness, and verbal working memory, despite scores similar to controls in measures of intellectual functioning. All dyslexics had non-word reading (Klepel test, van den Bos *et al.* (1994)) scores below percentile 5 compared to the university norm group (Depessemier and Andries, 2009), and had a formal diagnosis of developmental dyslexia. None of the controls reported any history of reading difficulties (Vandermosten *et al.*, 2010). Psychoacoustic measures of temporal auditory processing were performed in the same two groups of subjects and are also included in Table 4.2, as they might shed light on the results. These measures included a tone-in-noise detection task, frequency modulation (FM) detection at a 2-Hz FM-rate, and the just noticeable difference (JND) in frequency at 490 Hz. Details of the testing methods are given in Appendix 4.A.

4.2.3 Stimuli

Stimulus waveforms were generated in MATLAB® with a 48000-Hz sampling rate and 16-bit resolution in the following way:

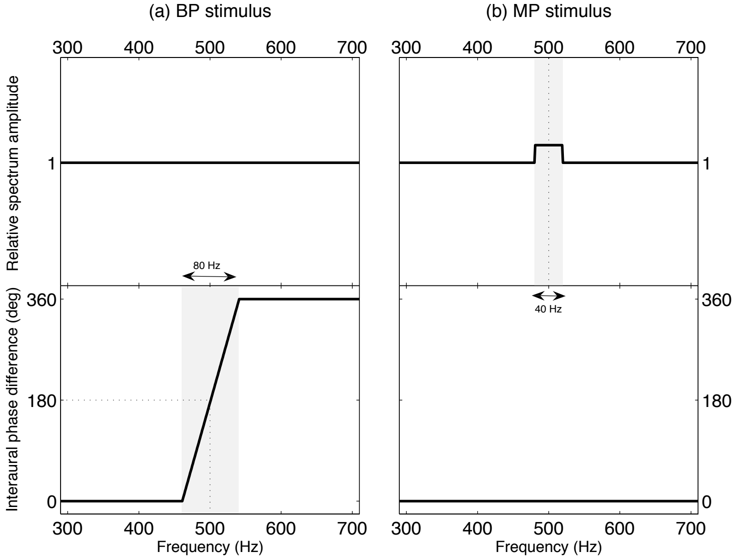


Figure 4.1: Amplitude and phase spectra of the two stimuli used in the pitch contour identification experiment. Example for a boundary frequency $f_b = 500$ Hz.

The BP stimulus was a Huggins' pitch, and contained a frequency-dependent interaural-phase-difference pattern, such that the left and right noises were in phase at all frequencies, except for a narrow frequency range around the boundary frequency f_b. In the transition area around f_b, a phase difference varying linearly from 0 to 2π was introduced in the frequency interval $[0.92f_b; 1.08f_b]$ (see Figure 4.1(a)), in order to create a pitch sensation corresponding to f_b. The stimuli were created as follows: (1) Random noise with the desired duration was generated in the spectral domain,

Table 4.2: Comparison of the cognitive and auditory profiles of: (a) control and dyslexic subjects; (b) dyslexic subjects from the D+ (normal pitch contour identification) and D- (deficit in pitch contour identification) groups. For each measure, the mean and standard deviation (in brackets) for both groups are given, as well as a p-value resulting from a two-sample t-test (paired for C/D, unpaired for D+/D-). Normality of the data was verified using a Shapiro-Wilk test. When normality was rejected at a 5% significance level (indicated by *) for at least one of the groups, a non-parametric Wilcoxon signed-rank test was used instead of a t-test. (c) Correlation of pitch identification scores with measures of cognitive function and temporal auditory processing. For each measure, Spearman's ρ and the corresponding p-value are given for each group. Total pitch identification scores with all stimulus configurations were used. For additional details about the auditory measures, see Appendix A.

	(a)					(b)					(c)			
					C/D					D+/D-	Controls		Dyslexics	
Measure	Controls		Dyslexics		p-value	Dyslexics D+		Dyslexics D-		p-value	p-val.	ρ	p-val.	ρ
Reading and spelling accuracy														
Reading real words (words/1 min.)	97.9	(10.5)	63.9	(10.6)	<.0001	64.1	(10.8)	63.6	(10.5)	.9089	.6154	(-.0939)	.2776	(-.2013)
Reading non-words (words/1 min.)	61.8	(10.0)	31.8	(5.8)	<.0001	32.4	(5.3)	30.4	(7.0)	.4103	.4889	(.1291)	.4686	(.1351)
Spelling real words (correct words)	25.0	(2.1)*	16.9	(3.7)	<.0001	16.5	(3.8)*	18.0	(3.2)	.2916	.5760	(.1045)	.2803	(-.2002)
Rapid automatized naming (RAN)														
RAN Colors (s)	26.2	(3.4)	32.7	(6.3)	<.0001	31.0	(3.9)	37.0	(8.8)	.0126	.5531	(-.1107)	.8107	(-.0448)
RAN Objects (s)	29.5	(2.9)	36.2	(5.6)	<.0001	35.6	(5.6)	37.8	(5.6)	.3235	.7422	(.0616)	.4021	(-.1560)
RAN Digits (s)	18.3	(3.1)*	24.7	(5.1)	<.0001	23.1	(4.3)	28.7	(4.7)	.0035	.5250	(-.1186)	.1232	(-.2828)
RAN Letters (s)	17.1	(2.8)	25.1	(5.8)	<.0001	23.4	(5.1)	29.2	(5.6)	.0087	.1842	(-.2449)	.2289	(-.2225)
Mean RAN reaction time (s)	22.8	(2.5)	29.7	(4.7)	<.0001	28.3	(3.8)	33.2	(4.9)	.0059	.4932	(-.1278)	.2542	(-.2111)
Phonemic awareness														
Phoneme deletion (% correct)	89.8	(9.0)*	81.9	(10.9)	.0024	83.7	(8.7)	77.6	(14.7)	.1578	.8770	(.0290)	.9112	(.0209)
Spoonerisms (% correct)	82.4	(10.4)	69.8	(14.9)*	.0006	83.1	(10.8)	80.7	(9.8)	.5713	.3297	(.1811)	.2762	(.2018)
Working memory														
Digit span (correct items)	14.3	(2.3)	11.6	(2.0)	<.0001	11.6	(2.0)	11.6	(2.0)	.9205	.5171	(.1209)	.5463	(.1126)
Non-word repetition (correct items)	25.6	(4.8)*	20.8	(4.8)	.0004	21.9	(4.8)	18.3	(3.9)	.0608	.0138	(.4376)	.2285	(.2227)
Intellectual functioning														
IQ score (Wechsler, 1999)	106.3	(9.6)*	107.8	(12.6)	.6034	110.5	(12.3)	101.2	(11.4)	.0636	.1637	(.2565)	.2069	(.2331)
Psychoacoustic performance														
Tone-in-noise detection (dB)	-9.4	(1.5)	-9.4	(1.6)	.8848	-9.5	(1.6)	-9.2	(1.6)	.6946	.8221	(.0421)	.6134	(-.0944)
Temporal auditory processing														
FM detection - 2 Hz (Hz)	4.7	(1.9)*	5.1	(2.3)*	.5435	4.9	(2.3)	5.4	(2.5)	.6332	.6196	(-.0928)	.7621	(-.0567)
Frequency JND - 490 Hz (%)	1.6	(1.7)*	1.9	(1.5)*	.3272	1.2	(0.9)*	3.8	(1.6)	.0001	.0521	(-.3521)	.0001	(-.6379)

using a 48000 Hz sampling rate; (2) All components of the noise were adjusted to have the same amplitude; (3) All frequency components above 4000 Hz were set to zero; (4) The stimulus obtained in step 3 was kept intact, transformed back to the time domain using the inverse Fourier transform (iFFT) algorithm, and fed to the left channel; (5) Some of the phase components of the stimulus obtained in step 3 were modified in order to create the desired interaural-phase-difference pattern, and the iFFT algorithm was applied to the stimulus.

The MP stimulus was generated in the same way as the BP stimulus, except that no interaural phase difference was introduced, *i.e.*, diotic broadband noise (BBN) was created. An additional diotic narrow band of noise (NBN) was then added to the BBN in the frequency interval $[0.96 f_b; 1.04 f_b]$, thereby creating a pitch sensation corresponding to f_b, due to an increased amplitude of frequency components around f_b (see Figure 4.1(b)). In order to obtain a similar salience for the MP and BP stimuli, the overall level of the NBN was adjusted using a linear relationship with the overall level of the BBN, following results from a preliminary salience adjustment experiment, described in Appendix 4.B. Although Huggins' pitch is generally lateralized towards one side of the head, the BP configuration used in this experiment was previously found to have an ambiguous lateralization towards either the left or the right side of the head (*cf.* HP^- in Raatgever and Bilsen, 1986). It was therefore chosen to introduce the NBN diotically rather than in one single channel, so that MP was perceived in the middle of the head.

For both stimulus types, each note was generated by adjusting f_b to the desired note frequency. Notes were then concatenated to form the different pitch contours, and each contour was preceeded and followed by 500 ms of diotic white noise (Figure 4.2). In order to avoid discontinuities in the waveform between successive notes, 1-ms onset and offset cosine ramps were used at the beginning and end of each portion of the stimulus. The overall stimulus was gated with 100-ms onset and offset cosine ramps. Wave files were created for each independent trial and implemented in the APEX 3 psychophysical platform (Francart *et al.*, 2008). Stimuli were fed through a LynxONE soundcard and presented at an overall level of 70 dB SPL *via* Sennheiser HDA 200 headphones in a sound-attenuating listening booth. Subjects were not informed about the existence of different stimulus types.

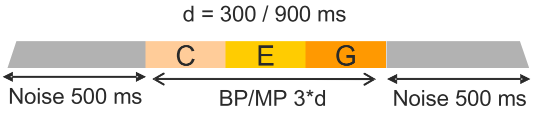


Figure 4.2: Stimulus design for the pitch contour identification experiment. Example of a rising pitch contour.

4.3 Results

4.3.1 Pitch contour identification experiment

Figure 4.3 shows the ability of control (light gray bars) and dyslexic (dark gray bars) subjects to *detect* the presence of pitch contours for each of the different stimulus configurations. It can be seen that subjects from both groups could clearly hear both MP and BP, independently of stimulus duration. In particular, the lowest overall score obtained among dyslexics with BP was 93%, showing that all dyslexic listeners without exception could hear binaural pitch. Differences between the two groups were overall not significant (MP300: p=.2344, MP900: p=.0156, BP300: p=.8594, BP900: p=.3594 [Wilcoxon signed-rank test]). N.B. None of the p-values mentioned in this chapter were corrected for multiple testing.

When comparing the total detection scores over trials with a 300-ms *vs.* a 900-ms note duration, no effect of stimulus duration was found on the ability of the listeners to *detect* the pitch contours (controls: p=.6719, dyslexics: p=.4688 [Wilcoxon signed-rank test]).

The average false-alarm rate, *i.e.*, the percentage of trials containing no pitch contour in which subjects pressed another button than the cross, was found to be rather low in both groups (dyslexics: 8.5%, controls: 6.8%), and never exceeded 35%. This

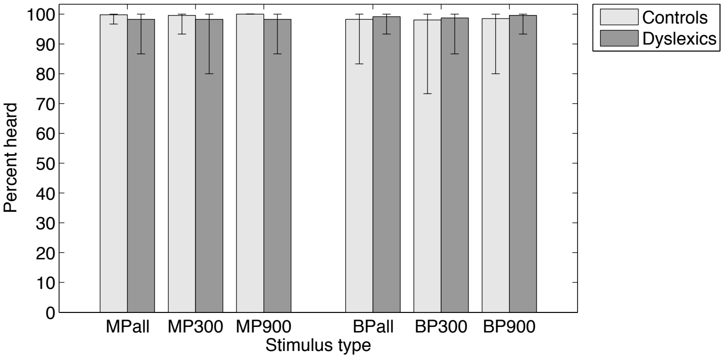


Figure 4.3: Percentage of trials containing a pitch contour in which the pitch contour was detected (*i.e.*, another button than the cross was pressed) for the different stimulus types. For MP and BP, the overall score regardless of duration is plotted first, followed by the scores for each duration independently (300 and 900 ms). Bar height corresponds to the mean over all subjects from a given group, and error-bars indicate the lowest and highest scores among all subjects from that group.

rules out the possibility that the high detection scores obtained here were due to strong false-alarm bias or a misunderstanding of the task.

If one now considers the ability of subjects to *correctly identify* the pitch contours (Figure 4.4), it appears that dyslexics are generally worse at the task than controls, in all stimulus configurations. This difference is only borderline significant when the whole group of dyslexic subjects is considered (MP+BP overall identification score: p=.0402 [Wilcoxon signed-rank test]). However, error-bars in Figure 4.4 indicate that the variability among subjects is higher in dyslexics than controls.

Figure 4.5(a) shows individual identification scores of MP and BP contours against each other. It can be seen that most dyslexic subjects actually performed similarly to controls (group D+, above the antidiagonal dashed line), while 9 dyslexics (group D-) and 1 control identified less than 80% of pitch contours correctly with both MP and BP (points with number-labels in Figure 4.5(a)), thus indicating difficulty with the task. The fact that all data points lie around the diagonal line in Figure 4.5(a) reflects that the stimulus type did not have an influence on the task, *i.e.*, the use of binaural pitch did not make pitch identification more difficult than for a monaurally-detectable pitch.

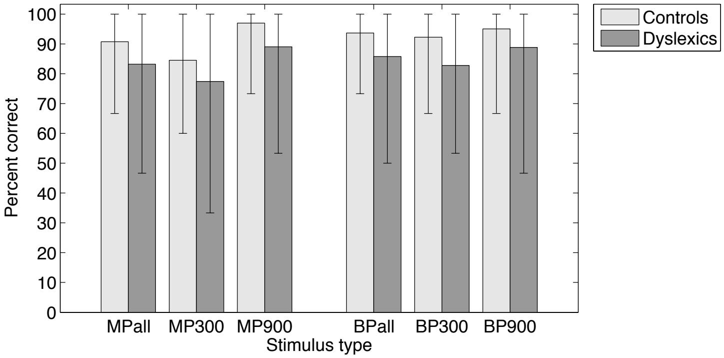
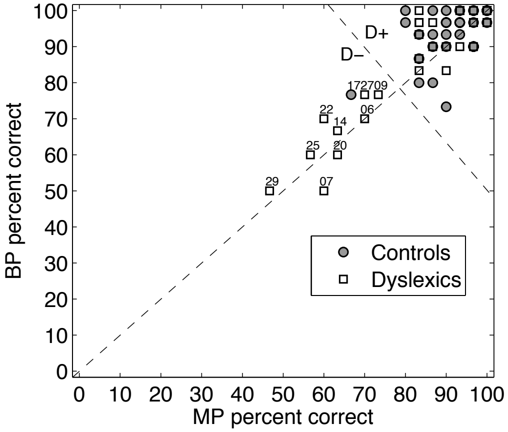


Figure 4.4: Percentage of trials containing a pitch contour in which the pitch contour was correctly identified, for the different stimulus types. For MP and BP, the overall score regardless of duration is plotted first, followed by the scores for each duration independently (300 and 900 ms). Bar height corresponds to the mean over all subjects from a given group, and error-bars indicate the lowest and highest scores among all subjects from that group.

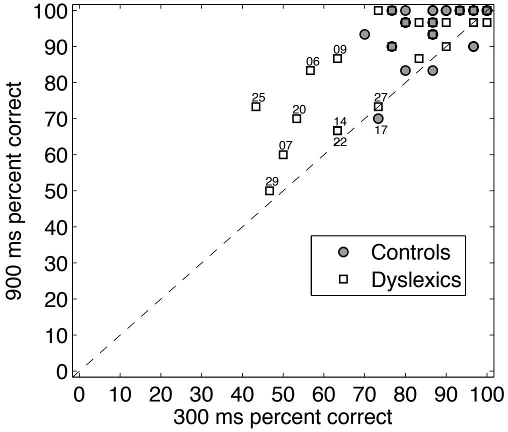
In fact, overall scores were on average higher with BP than MP. Moreover, 7 of the 10 labelled subjects in Figure 4.5(a) also obtained less than 80% correct identification in the practice run with pure-tone stimuli (only subjects 9, 14 and 27 obtained more than 80% correct in that condition), suggesting that their difficulty stems from the nature of the task rather than the type of stimulus used.

When comparing overall identification scores obtained with short *vs.* long note durations (Figure 4.5(b)), it appears that almost all subjects benefited from a longer note duration (points generally fall above the diagonal line). Average scores for MP and BP stimuli were found to be significantly higher with 900-ms notes than 300-ms notes in both groups of listeners (dyslexics: p<0.0001, controls: p=0.0001 [Wilcoxon signed-rank test]). The analysis of recorded reaction times revealed no significant difference between dyslexics and controls.

The analysis of error matrices showed that more misses occurred for the constant pitch contour than for the rising and falling pitch contours, and that the rising and falling pitch contours were confused with each other more often than with the constant



(a) Stimulus type



(b) Note duration

Figure 4.5: Influence of stimulus type and note duration on the ability to correctly identify pitch contours: Percentage of trials containing a pitch contour in which the pitch contour was correctly identified: (a) with MP (horizontal axis) *vs.* BP (vertical axis) stimuli; (b) with 300-ms (horizontal axis) *vs.* 900-ms (vertical axis) note durations.

pitch contour. These trends were, however, similar in both the dyslexic and control groups.

4.3.2 Correlation with measures of cognitive function and temporal auditory processing

Overall, no significant correlations were found between pitch identification scores and cognitive measures included in the subjects' profile (Table 4.2(c)), *i.e.*, measures of reading and spelling accuracy, rapid automatized naming (RAN), phonemic awareness, working memory, and intellectual functioning. Taking multiple statistical testing into consideration, the only psychoacoustic measure that correlated significantly to performance in pitch contour identification was frequency discrimination in the dyslexic group. Performing the correlation analysis on the total pool of subjects (dyslexics and controls) led to even higher significance of the correlation between pitch contour identification scores and all measured frequency JNDs (p<.0001 [ρ=-.4960]). A scatter plot of overall pitch contour identification scores *vs.* the measured frequency JNDs is given in Figure 4.6.

When comparing the cognitive and auditory profiles of dyslexic subjects from groups D+ and D- (see section 4.3.1), it appears that both groups show similar performance in most tasks (Table 4.2(b)). However, a significant group difference was found for frequency discrimination abilities as well as RAN response times for digits, letters, colors, and mean RAN response times. It is also worth noting that the only control subject who had difficulty with pitch contour identification (subject 17) performed poorer than all other controls in several tasks (FM detection, spelling of real words) and obtained low scores in the phoneme deletion and digit span tests, despite similar reading scores to other controls. The latter subject also showed frequency JNDs that were overall considerably higher than in other controls.

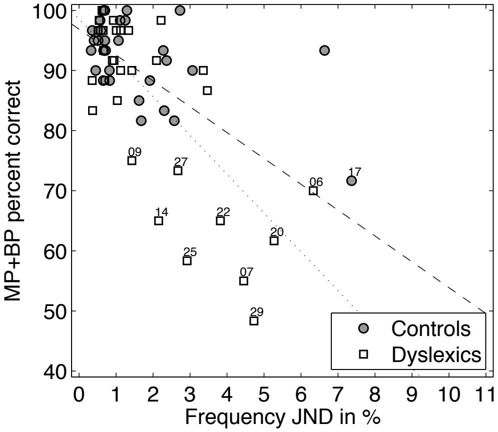


Figure 4.6: Correlation between results in pitch contour identification and frequency discrimination. Total pitch identification scores with all stimulus configurations are plotted against the just noticeable difference (JND) in frequency at 490 Hz. Straight lines show the correlation obtained over the whole pool of subjects (dashed line), and over the dyslexic group alone (dotted line).

4.4 Discussion

The present results clearly show that dyslexic listeners are able to perceive binaural pitch. Moreover, the pitch detection scores of subjects with dyslexia were similar for the BP and MP stimuli. Because perceiving BP requires the comparison of accurate phase information across ears, these two findings suggest that no severe dysfunctions in peripheral temporal fine structure processing or binaural integration mechanisms are associated with dyslexia, and confirm the findings of Chait *et al.* (2007), who found no sign of binaural impairment in dyslexia using binaural pitch stimuli. This conclusion is supported by the fact that FM detection scores of the subjects with dyslexia suggested normal temporal fine structure processing (Moore and Sęk, 1996) in dyslexic listeners. These normal FM detection scores contrast with the reduced FM detection abilities previously found in pre-school children (Boets *et al.*, 2007), and might reflect the presence of compensation mechanisms in these adult dyslexic subjects (university students).

The fact that all subjects could easily detect BP, but that a subgroup of the dyslexics (29%) had difficulty with pitch contour identification for both MP and BP stimuli, suggests that the findings of Dougherty *et al.* (1998) and Edwards *et al.* (2004) may have been task-related: They used pitch contour identification and lateralization tasks, while Chait *et al.* (2007) used a simple detection task. Another difference that one should bear in mind is that the former two studies tested children, while the latter used adult subjects, who might have developed compensation mechanisms and thus show higher performance. However, it is unlikely that compensation can explain the whole of the present findings concerning binaural pitch perception: The presence of a subgroup with reduced performance in the present study confirms that the task remains problematic for these adult subjects. Moreover, because no influence of note duration on detection scores was found, it is unlikely that the results obtained by the former two studies were due to short stimulus durations. This is in line with findings from Banai and Ahissar (2006), who showed that the psychoacoustic abilities of dyslexic listeners with additional learning difficulties depended on the complexity of the required task rather than the nature of the presented stimuli. In the present study, the lack of significant difference between reaction times of dyslexics and controls may also reflect an influence of task complexity: While the present task required decision between four response buttons, leading to long response times in both groups of subjects, Chait *et al.* (2007) obtained significantly longer response times in dyslexics than controls with a more automatic task.

One question raised by the present results concerns the origin of the difficulty of the D- group with pitch contour identification. Given the nature of the task and the experimental paradigm used in this study, several suggestions can be made that might explain this difficulty.

One explanation could be that dyslexics of the D- group have a difficulty detecting tonal objects in background noise, as suggested by Chait *et al.* (2007). This would mean that, for these subjects, MP and BP are less salient than for the control and D+ groups, making pitch contour identification more difficult when using such stimuli. However, most D- subjects also had difficulty with the task in the training session with pure tones, which contained no background noise. Additionally, no significant group difference between D+ and D- subjects was found in the tone-in-noise detection task. Therefore, it is unlikely that a weaker pitch sensation with MP and BP stimuli was responsible for the lower pitch identification scores in the D- group.

Another explanation could be that, in the experimental procedure, subjects had to link an auditory pattern to a visual symbol: Each possible pitch contour corresponded to a different response button and, for instance, subjects had to link a rising pitch contour to an upward-pointing arrow. Such an ability might be impaired in some dyslexics. If such a deficit was the main reason for low pitch contour identification scores in the D- group, one would expect D- subjects to perform as D+ subjects in a similar task that does not involve linking an auditory pattern to a visual symbol. In the present study, the frequency discrimination task was close to such a situation: While a comparison of the pitch of successive tones was required, no visual symbols had to be linked to specific pitch directions in the response process, as the subjects were instructed to identify the odd interval. However, a significant group difference in frequency JND was still found between D+ and D- subjects. This suggests that a difficulty linking auditory and visual patterns does not satisfactorily explain the results of the D- group.

Despite the rather large intervals between successive notes used in this study, one cannot exclude the possibility that impaired frequency discrimination was responsible for making pitch contour identification more difficult in the D- group. One might argue that the significant correlations between pitch contour identification scores and frequency JNDs support this hypothesis. However, frequency JNDs at 490 Hz in the D- group never exceeded 8% of the test frequency. Because frequency intervals in the pitch contours used here were larger than 17%, it can be assumed that all subjects were able to discriminate between successive notes.

The question remains why some subjects failed to identify the individual contours even though they could hear the difference between them, and why frequency JNDs are then correlated to pitch contour identification scores. When comparing the subjects' tasks in the pitch contour identification and the frequency discrimination experiments, one can observe that they are, in fact, very similar: Frequency discrimination was measured using a three-alternative forced-choice (3AFC) paradigm, in which subjects listened to three successive tones before deciding which of these tones had a different pitch than the other two. This corresponds to choosing between three possible pitch contours, and might explain why results from the two experiments strongly correlate.

This raises the question whether frequency discrimination *per se* is really impaired in some dyslexic listeners, or whether the obtained results just reflect a limit imposed by a difficulty with the nature of the task itself. The present study used a three-interval, three-alternative forced-choice (3I-3AFC) paradigm and found higher mean frequency JNDs in the dyslexic group than in the control group, but these group differences were not or only borderline significant. McAnally and Stein (1996) found a strongly significant difference between frequency JNDs of dyslexics and controls at 1 kHz, using a two-interval same-different paradigm in which the reference was presented once (2I-1A-X). Hill *et al.* (1999) measured frequency JNDs at 1 and 6 kHz with a 4I-2AFC paradigm in which the second or the third interval contained the target and found no group difference between dyslexics and controls at either test frequency. Considering such different results obtained using different tasks, it appears essential to investigate and discuss the influence of the experimental procedure on frequency JNDs with dyslexic listeners. In their comparison of thresholds obtained with a 2I-1A-X paradigm and a 2I-6A-X paradigm in which the reference was presented six successive times, France *et al.* (2002) showed that JNDs of dyslexic listeners could be reduced to those of controls by increasing the number of available observations and using short inter-stimulus intervals. They suggested that a deficit in early auditory memory (Hari *et al.*, 1999) could explain the dependence of JNDs on the procedure used, and argued that repeated exposure to known identical references might help stabilize auditory memory and thus lead to lower thresholds.

If the difficulty of D- subjects in identifying pitch contours disappears when changing the experimental procedure, this would confirm that these subjects are in fact able to perceive the difference between successive stimulus intervals, and that their difficulties are directly linked to the nature of the task. Therefore, mechanisms responsible for the ability to retain successive stimulus intervals in memory could be deficient. This would be consistent with the presence of a significant D+/D- group difference in a discrimination task (frequency JND), but not in less-complex detection tasks (tone-in-noise and FM detection). This hypothesis would also be in line with findings from Banai and Ahissar (2004): A subgroup of their dyslexic subjects (DP) obtained frequency JNDs that were elevated compared to other dyslexics, and subjects from this subgroup, which formed a proportion similar to that of the D- group of the present study, were also the ones showing a significant impairment in verbal working memory. Moreover, in both studies, significantly longer RAN reaction times were found in subjects from the aforementioned subgroups (DP in Banai and Ahissar (2004), D- in the present study), compared to other dyslexic subjects. This suggests that the core phonological deficit of D+ and D- subjects might have different etiologies, and that a deficit in rapid information retrieval from memory plays a role in the difficulty of D- subjects with the pitch contour identification and frequency discrimination tasks.

Finally, because pitch contour identification involves following changes in pitch, the ability to switch attention from one pitch percept to the next could also be impaired in the D- group. This would be consistent with findings from Hari and Renvall (2001) whose results suggested that "sluggish attentional shifting" could give rise to impaired processing of rapid stimulus sequences. More recently, Hämäläinen *et al.* (2008) measured event-related potentials (ERP) in reading-disabled children and found that ERP responses to pitch changes were lower in reading-disabled children than control children, in a component related to attention switching.

4.5 Conclusion

It was found that binaural pitch was easily detectable in both dyslexic listeners and matched controls, which suggests intact low-level binaural processing in dyslexia. In both groups of subjects, pitch contour identification scores were similar for binaural pitch stimuli and monaurally-detectable pitches in noise, showing no sign of low-level binaural impairment in dyslexic listeners. A subgroup of dyslexics showed difficulties with pitch contour identification. Results in that experiment were significantly correlated with measures of frequency discrimination, and this correlation is most likely due to the similarity of the tasks in the two experiments. The difficulty in such tasks is thought to be attributable to auditory memory or auditory attention deficits, rather than reduced frequency discrimination abilities *per se*. The results favor impaired cognitive mechanisms as precursors to reading disability, rather than impaired low-level auditory processing. They underline the influence of the choice of the experimental paradigm and the task of the subjects on results from basic psychophysical measures with dyslexic listeners. Overall, great care ought to be taken before asserting the presence of a low-level auditory processing deficit in a dyslexic group, if the task involves auditory memory or auditory attention to a non-negligible extent.

Appendix 4

A. Methods used for auditory processing measures

A 3AFC procedure was used in all experiments. *Psychoacoustic performance:* Detection of a 1-kHz pure tone in background noise was measured. Results are given as the signal-to-noise ratio at threshold in dB. *FM detection:* Detection of frequency modulation of a 1-kHz tone was measured for a 2-Hz FM-rate. Results are given as the maximum frequency excursion at threshold in Hz. *Frequency discrimination:* The just noticeable difference in frequency was measured at 490 Hz using a 2-down 1-up procedure and a fixed-reference paradigm. Targets were always lower in frequency than the reference tone and the target frequency was varied by a factor of 1.4. Stimuli were presented monaurally at 70 dB SPL. Results are given as the smallest detectable change in frequency (in % of the test frequency).

B. Salience adjustment of MP and BP

In order to match the salience of MP to that of BP, a preliminary salience adjustment experiment was performed by 5 normal-hearing listeners. A 2I-2AFC procedure was used, in which one random interval contained a BP stimulus, and the other interval contained an MP stimulus. The overall level of the broadband noise in both MP and BP stimuli (L_{BBN}) was fixed, and the tracking variable was the overall level of the additional narrow band of noise in the MP stimulus (L_{NBN}). For each presentation, the task of the listener was to indicate, *via* a computer interface, in which interval the pitch was more salient. Intervals had a 500-ms duration, including 30-ms onset and offset cosine-ramps, and were separated by a 500-ms silent pause. A 1-up 1-down procedure was used: when BP was perceived as more salient, L_{NBN} was increased in the next presentation, and when MP was perceived as more salient, L_{NBN} was decreased in the next presentation. The starting value of L_{NBN} was 75 dB SPL. Stepsizes of 8, 4, 2, and 1 dB were used, and the stepsize was decreased after each upper reversal. A run was terminated after 14 reversals and the threshold value was determined from all points following the 6$^{\text{th}}$ reversal. The experiment was performed for $L_{\text{BBN}} = [55;60;65;70;75;80]$, with $f_b = 500$ Hz. Stimuli were generated as described in section 4.2.3. Each subject performed three runs for each value of L_{BBN}. The best-matching L_{NBN} value was defined as the average value obtained over all runs. The average L_{NBN} giving equal salience was found to be linearly correlated to L_{BBN}, and a first-degree polynomial, described by $L_{\text{NBN}} = 1.07 \times L_{\text{BBN}} - 15.69$, was fitted to the data. This relationship was used to generate the MP stimulus for the pitch contour identification experiment (see section 4.2.3).

Acknowledgments

This work was carried out at the Division of Experimental Otorhinolaryngology (ExpORL), Department of Neurosciences, Group Biomedical Sciences, Katholieke Universiteit Leuven, Belgium. Many thanks to Hanne Poelmans and Heleen Luts for carrying out the pitch contour identification experiment and for their great help throughout the whole project, as well as Jan Wouters and all contributors to the DYSCO project for their time and fruitful discussions.

5

The role of temporal fine structure information for the low pitch of high-frequency complex tones[§]

The fused low pitch evoked by complex tones containing only unresolved high-frequency components demonstrates the ability of the human auditory system to extract pitch using a temporal mechanism in the absence of spectral cues. However, the temporal features used by such a mechanism have been a matter of debate. For stimuli with components lying exclusively in high-frequency spectral regions, the slowly varying temporal envelope of sounds is often assumed to be the only information contained in auditory temporal representations, and it has remained controversial to what extent the fast amplitude fluctuations, or temporal fine structure (TFS), of the conveyed signal can be processed. Using a pitch-matching paradigm, the present study found that the low pitch of inharmonic transposed tones with unresolved components was consistent with the timing between the most prominent TFS maxima in their waveforms, rather than envelope maxima. Moreover, envelope cues did not take over as the absolute frequency or rank of the lowest component was raised and TFS cues thus became less effective. Instead, the low pitch became less salient. This suggests that complex pitch perception does not rely on envelope coding as such, and that TFS representation might persist at higher frequencies than previously thought.

[§] This chapter is based on Santurette and Dau (2011a).

5.1 Introduction

Pitch is the attribute of auditory sensation allowing us to order sounds on a musical scale (ASA, 1960). It is used as a cue for human perception of a variety of ordinary sounds, and contributes to important elements of hearing such as music perception, speech recognition, and sound source segregation (Plack and Oxenham, 2005b; Bregman, 1990). Despite this extensive use of pitch by the auditory system for the perception of everyday sounds, the way pitch extraction mechanisms make use of the spectral and temporal features of sounds remains poorly understood. The human auditory system is able to process sounds with both a fine spectral and temporal resolution, and the debate of whether spectral or temporal cues are dominant for pitch retrieval is central to pitch perception research (de Cheveigné, 2005).

In one classical view, the basilar membrane in the cochlea acts as a frequency analyzer, and the resulting spectral information is preserved along the auditory pathway due to its tonotopic organization (Rose *et al.*, 1959; Merzenich *et al.*, 1975). Moreover, most ordinary sounds can be analyzed as a sum of sinusoidal components with harmonic frequencies and evoke a pitch corresponding to their fundamental frequency (F0). Auditory mechanisms may therefore retrieve the pitch of sounds by matching harmonic templates to a spectral excitation pattern consisting of the characteristic frequencies of maximally-stimulated auditory channels. Such rate-place mechanisms are based solely on the tonotopic property of the auditory system, and are the basis for spectral pattern-matching models of pitch perception (Wightman, 1973; Terhardt, 1974; Cohen *et al.*, 1995).

In an alternative view, vibration of the basilar membrane gives rise to synchronous firing of auditory-nerve cells (Rose *et al.*, 1967). This implies that precise temporal information is also available as a tool for pitch extraction. Therefore, most temporal models of pitch perception exclusively make use of the distribution of inter-spike intervals (ISI) in the auditory nerve (Licklider, 1951; de Cheveigné, 1998; Meddis and Hewitt, 1991) to retrieve pitch. At low frequencies, ISIs reflect intervals between maxima in the temporal fine structure (TFS), *i.e.*, the fast amplitude fluctuations, of the sound waveform. If components of the sound are separated (resolved) on the basilar membrane, the most frequent ISIs can thus represent the period of individual frequency components in the corresponding channels. If neighboring components interact on the basilar membrane, ISIs may reflect the timing between TFS peaks close to adjacent envelope maxima. As frequency increases, the ability of auditory-nerve cells to phase lock and follow the fast waveform fluctuations is progressively reduced. Therefore, the accuracy with which ISIs represent TFS-related intervals is expected to decay, causing the TFS-related peaks in an ISI histogram to merge into an envelope-related maximum. Thus, it is generally assumed that only the temporal envelope of the sound waveform, *i.e.*, its slow amplitude variations, is available as a temporal cue at high frequencies.

Pattern-matching models, which make use of information concerning either the place of excitation or the periodicity of individual frequency components of the sound (Walliser, 1969; Goldstein, 1973), require these components to be spectrally resolved on the basilar membrane, such that distinct peaks of excitation on the tonotopic axis occur, or ISIs related to individual component periods are produced. Temporal models based on ISI histograms, however, do not suffer from such a limitation and can in principle account for the pitch of sounds containing only higher, spectrally-unresolved components. Although most pitch-evoking sounds in our environment contain resolved harmonics, the low pitch evoked by complex tones with unresolved components has raised particular interest in the literature (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Hall *et al.*, 2003). In the absence of spectral cues, pitch mechanisms must rely on the temporal features available to them after cochlear processing. Such features may include information about the stimulus envelope as well as information about its TFS.

The relative importance of envelope and TFS information for the low pitch of high-frequency complex tones has long been a matter of debate. For instance, a small change in the carrier frequency of an amplitude-modulated sine wave produces a shift in its low pitch (de Boer, 1956a). As this change in carrier frequency does not affect the envelope repetition rate of the stimulus, but modifies its TFS, it has been suggested that the corresponding pitch shift may reflect the use of TFS information (de Boer, 1956b; Schouten *et al.*, 1962). Following this idea, recent studies have shown that normal-hearing listeners are able to use a pitch cue to discriminate between harmonic and frequency-shifted (inharmonic) complex tones, band-pass filtered in a region where all their components were unresolved (Moore *et al.*, 2009b; Moore and Sęk, 2009a). This challenged theories according to which resolvability of some components is required to produce a pitch shift between a harmonic and a frequency-shifted complex. Furthermore, discrimination remained possible even when all audible components of the complex were located above 8 kHz (Moore and Sęk, 2009a). This suggested that TFS information remains available to the human auditory system and contributes to pitch perception in frequency regions higher than the traditional frequency limit above which accurate representation of TFS is generally assumed to vanish. A value of 5 kHz has often been proposed for such a limit, which is consistent with a loss of accurate frequency discrimination (Moore, 1973) and melody perception (Ward, 1954; Attneave and Olson, 1971) above this frequency. However, the upper frequency limit for observable phase-locked activity in the auditory pathway is highly species dependent (Johnson, 1980; Palmer and Russell, 1986; Köppl, 1997) and remains unknown in humans.

The recent findings of Moore *et al.* (2009b) and Moore and Sęk (2009a) suggested a crucial role of TFS in extracting the low pitch of high-frequency complex tones with unresolved components. However, it remains controversial whether the pitch difference between the harmonic stimuli and their frequency-shifted counterparts, which is assumed to be the main cue in their discrimination task, relied on TFS or spectrally-conveyed information. First, shifting partials of a complex introduces changes in the shape of the internal excitation pattern it produces, which might be detectable by the listeners (Moore and Moore, 2003). Second, combination tones arising from non-linearities in the cochlea shift in frequency when the physical components of the complex are themselves shifted and, if audible, might provide additional cues (Oxenham *et al.*, 2009). Despite efforts made to minimize differences in excitation pattern shape and to mask audible combination tones in the aforementioned studies (Moore *et al.*, 2009b; Moore and Sęk, 2009a), it has been questioned whether these potential spectral indices were totally eliminated (Oxenham *et al.*, 2009). Furthermore, it remains possible that temporal envelope cues at the output of auditory filters might play a role in discriminating harmonic and frequency-shifted stimuli.

In the present study, a pitch matching paradigm with a single reference stimulus was used. In other words, the pitch of each reference stimulus under investigation was matched against that of a broadband harmonic complex with variable F0. Unlike a discrimination task, this method prevented the use of changes in spectral indices as a cue, and had the additional advantage of producing a pitch estimation. Transposed tones (van de Par and Kohlrausch, 1997) with inharmonic, unresolved components were used as reference stimuli. Such stimuli were chosen because they are traditionally assumed to convey only temporal envelope cues at the level of the auditory nerve, when their components lie in high spectral regions (Oxenham *et al.*, 2004). Moreover, they have similar power spectra to the complex tones used by Moore *et al.* (2009b), allowing a comparison of the pitch matches to their findings. The transposed stimuli were designed such that the time intervals between successive maxima in the envelope waveform differed from the time intervals between the most prominent peaks in the TFS waveform (Fig. 5.1). This made it possible to investigate whether the resulting low pitch corresponded to the inverse of the envelope repetition rate f_{env} (Fig. 5.2, Hypothesis A: unimodal distribution of pitch matches around f_{env}), the inverse of the timing between their most prominent TFS peaks (Fig. 5.2, Hypothesis B: multimodal distribution of matches around subharmonics of the carrier frequency f_c), or whether they do not evoke a salient pitch (Fig. 5.2, Hypothesis C: random matches). The aim was to determine whether the low pitch of such high-frequency stimuli relies preferably on TFS cues, on envelope cues, or whether none of these cues are sufficient for pitch extraction, depending on the frequency region considered.

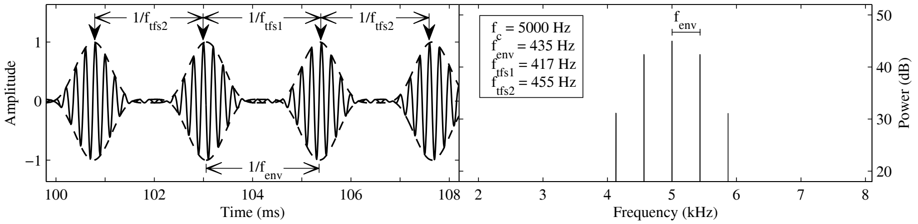


Figure 5.1: Temporal waveform (left panel, solid line), envelope (left panel, dashed line) and power spectrum (right panel) of a transposed tone with a carrier frequency f_c=5000 Hz and an envelope repetition rate f_{env}=435 Hz. In such an example, where the ratio $N=f_c/f_{env}$ is not an integer, the timing between the most prominent fine structure peaks (down-pointing arrows) in two successive envelope periods equals one of two values ($1/f_{tfs1}$ or $1/f_{tfs2}$), which always differ from the envelope period $1/f_{env}$.

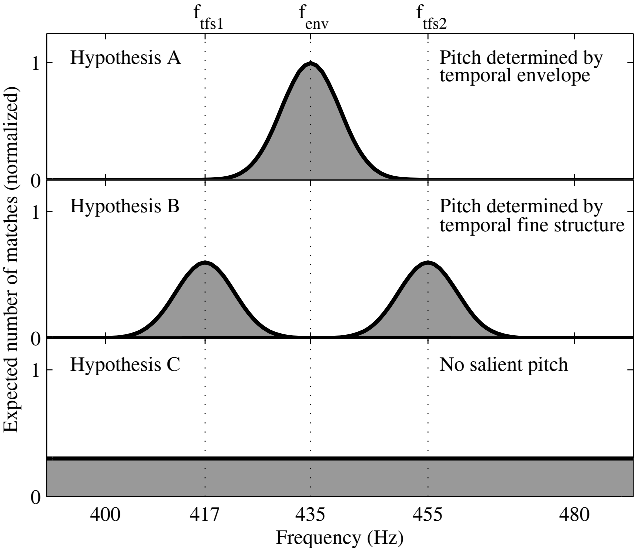


Figure 5.2: Expected distributions of pitch matches for each of the three experimental hypotheses. A: Transposed tones evoke a pitch corresponding to their envelope repetition rate. B: Transposed tones evoke an ambiguous pitch related to the fine timing of most prominent peaks in their carrier. C: Transposed tones do not evoke a salient pitch. Given frequencies correspond to the example from Fig. 5.1 (f_c=5000 Hz and f_{env}=435 Hz).

5.2 Methods

5.2.1 Stimuli

The transposed tones were generated by multiplying a carrier pure tone with frequency f_c with a half-wave rectified sinusoid with frequency f_{env}. Before multiplication, the half-wave rectified sinusoid was low-pass filtered using a 256-tap FIR filter (cut-off frequency $0.2f_c$) designed after a fourth-order Butterworth response. Respective carrier frequencies and envelope repetition rates (f_c/f_{env}), in Hz, of 3000/261, 3000/207, 4000/348, 4000/276, 5000/435, 5000/345, 6000/522, 6000/414, 7000/609 and 7000/483 were used, for a total of 10 experimental conditions. Ratios between f_c and f_{env} of $N \approx 11.5$ and $N \approx 14.5$ were chosen, so that f_c was either located between harmonics 11 and 12 or 14 and 15 of f_{env}. Such ratios gave rise to stimuli with envelope periods that differed clearly from the timing between the most prominent TFS maxima. Moreover, as the limit for harmonic resolvability is usually found to lie somewhere between harmonics 5 and 10 (Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; Moore and Ohgushi, 1993; Plomp, 1964; Plomp and Mimpen, 1968), it could be assumed that all frequency components of the stimuli were unresolved. An additional condition was also included in which components of the transposed tone were in a harmonic relationship (f_c=3000 Hz, f_{env}=250 Hz, N=12).

The pitch of broadband pulse trains (Fig. 5.3) was matched to that of the transposed tones. The choice of pulse trains as matching stimuli was motivated by the fact that they have a closer timbre to that of transposed tones than regular pure tones, while they still produce a clear pitch sensation. In addition to facilitating pitch comparisons, this was assumed to make timbre differences less likely to interfere with pitch judgements (Moore *et al.*, 1992). Pulse trains with a period of $1/f_p$ were generated by adding harmonic cosine tones of f_p, starting at the fifth harmonic. The resulting pulse train was passed through a 512-tap FIR filter (cut-off frequencies 2 and 10 kHz) designed after a fourth-order band-pass Butterworth response.

The overall level of the stimuli was 50 dB SPL for transposed tones and 55 dB SPL for pulse trains, and the duration of all stimuli was 500 ms, including 30-ms onset

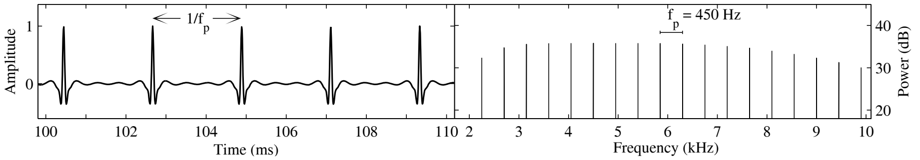


Figure 5.3: Temporal waveform (left panel) and power spectrum (right panel) of a broadband pulse train with a fundamental frequency f_p=450 Hz, evoking a pitch at f_p.

and offset cosine ramps. In order to mask combination tones, pink noise, band-pass-filtered from 100 to 12000 Hz (512-tap FIR filter designed after a fourth-order Butterworth response), was added to the stimuli at a spectrum level of 13.5 dB/Hz at 1 kHz (decay 3.01 dB per octave). Such a noise level corresponded to about 34 dB SPL per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) in the frequency region of the stimuli, which was similar to the relative noise levels used in earlier studies (Moore *et al.*, 2009b; Moore and Sęk, 2009a; Oxenham *et al.*, 2009). Given this pink-noise level and the low-pass filtering of the half-wave rectified modulator used to generate the transposed tones, the latter always contained 5 effective components ranging from f_c-2f_{env} to f_c+2f_{env} (*cf.* Fig. 5.1). Moreover, as combination tone levels in this frequency range are expected to lie at least 10 dB below the level of the lowest component (Zwicker, 1981), and because the spectrum level of pink noise increases with decreasing frequency, it was assumed that all combination tones were properly masked. All stimuli were generated in MATLAB and presented with a 96-kHz sampling rate and a 32-bit resolution *via* an RME DIGI96/8 soundcard and Sennheiser HD580 calibrated headphones in a double-walled sound-attenuating listening booth.

5.2.2 Experimental procedure

Through a computer interface, the listeners were able to play the reference transposed tone and target pulse train as many times as they wished, until they were satisfied with the match. The background pink noise was played continuously throughout the matching procedure. The F0 of the pulse train, f_p, could be varied in steps of

4 semitones, 1 semitone, or 1/4 semitone. Subjects were encouraged to use larger steps at first, and reduce them progressively. In order to avoid octave confusions, a visual cue appeared on the screen when f_p deviated from f_{env} by more than 2/3 octave. The starting value of f_p for each presentation was randomly chosen from a uniform distribution of values between $0.8f_{env}$ and $1.2f_{env}$. A total of 50 matches per condition were obtained for each listener, of which the last 40 were included in the results. The experiment was divided into 10 runs of 55 matches each and, in each run, the presentation order of the different experimental conditions was randomized.

5.2.3 Subjects

Six normal-hearing test subjects (3 female, 3 male, ages: 23-32 years, mean age: 25.7 years) participated in the experiment, which was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-KA-04149-g). All subjects had hearing thresholds below 20 dB HL at all audiometric frequencies in both ears. They were tested monaurally in their best ear, defined as the ear with the lowest average hearing threshold between 2 and 8 kHz. The mean hearing threshold over all tested ears was -0.5 dB HL. All subjects had some form of musical training and played an instrument as a hobby.

5.2.4 Preliminary experiment

The pitch accuracy of the subjects was tested in a preliminary experiment, in which they were asked to match the pitch of pure tones to that of broadband pulse trains, similar to those used in the main experiment. In addition to familiarizing the subjects to the experimental procedure, this first test was used to confirm that such pulse trains elicited an unambiguous low pitch, and could thus be used as reliable matching stimuli. The procedure was the same as described above, and 10 different pulse trains, with f_p values equal to the different f_{env} values used in the main experiment, were used as reference stimuli. The frequency of a sinusoidal tone at 60 dB SPL was the adjustable variable, and the pink-noise background was also present in this experiment. Ten subjects participated in this preliminary experiment. Each subject performed a minimum of 6 runs of 50 matches each, until their pitch accuracy remained stable.

In all conditions, all subjects were able to hear a low pitch, with median matches close to f_p. However, standard deviations differed across listeners and some showed a higher number of erratic matches than others. A Grubbs' test for outliers (Grubbs, 1969) was performed on the last 20 matches obtained for each condition. On this basis, four of the participants, whose overall dataset contained more than 5% outlying matches, were excluded from the main experiment. This ensured similar pitch accuracy across subjects, and increased the likelihood of obtaining distributions of matches with sufficiently small variances to distinguish between the different hypotheses of Fig. 5.2. For the six remaining subjects, the average pitch accuracy, measured as the average of group standard deviations over all conditions (third row in Table 5.1), was 2.8 Hz. This confirms that the broadband pulse trains evoked a sufficiently precise low pitch to be used as matching stimuli. Overall, the pulse trains were found to have a lower pitch than a pure tone with frequency f_p, by a factor of 0.986 on average (Table 5.1), which is consistent with pitch shifts previously found between pure tones and complex tones (Moore *et al.*, 1992).

Table 5.1: Mean pure-tone frequencies for a pitch match with a broadband pulse train with a fundamental frequency f_p and mean of individual standard deviations over the six subjects with the highest pitch accuracy.

f_p (Hz)	207	261	276	345	348	414	435	483	522	609
Pure tone match (Hz)	201.2	255.5	272.2	341.3	343.6	411.3	429.6	477.8	515.3	604.9
Ratio to f_p	0.972	0.979	0.986	0.989	0.987	0.993	0.988	0.989	0.987	0.993
Mean standard deviation (Hz)	3.4	2.2	1.8	2.2	2.1	2.4	2.5	3.0	3.6	4.7

5.3 Results

5.3.1 Pitch matching

Histograms were built from the raw data of the whole subject group (240 matches per condition), using a bin width of $f_{env}/250$ (normalized to the harmonic condition), which always corresponded to about 7% of a semitone. The distributions of pitch matches obtained for each condition are shown in Fig. 5.4, 5.5, and 5.6.

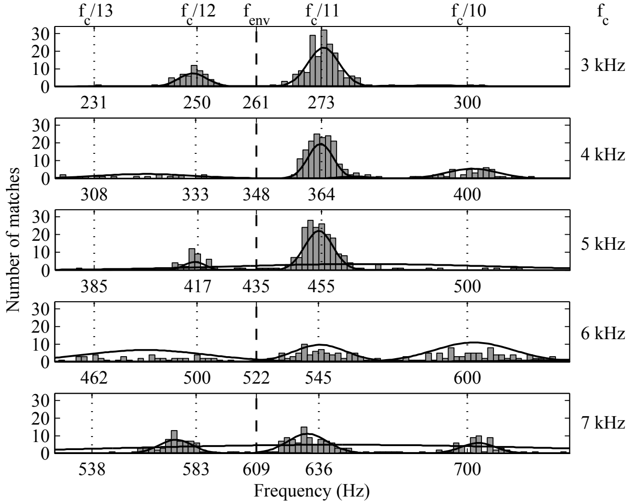


Figure 5.4: Pitch matching of the fundamental frequency of broadband pulse trains (horizontal axis) to transposed tones with a carrier frequency f_c and an envelope repetition rate f_{env}, with $N=f_c/f_{env} \approx 11.5$. The total distribution of pitch matches for all six subjects is shown (histograms and fitted Gaussian mixture models), with 40 matches per condition per subject. The vertical dashed lines indicate f_{env} for each condition, while the dotted lines indicate the frequencies corresponding to the inverse of time intervals between the most prominent TFS maxima. f_c values are indicated to the right of each plot.

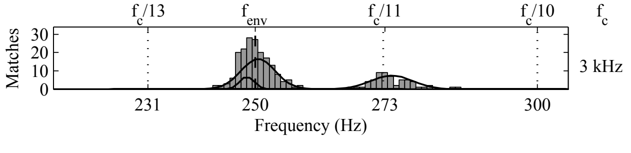


Figure 5.5: Pitch matching of the fundamental frequency of broadband pulse trains (horizontal axis) to transposed tones with a carrier frequency f_c and an envelope repetition rate f_{env}, with $N=f_c/f_{env}=12$. See Fig. 5.4 caption for more details.

For N≈11.5 (Fig. 5.4), multimodal distributions were obtained and, overall, the pitch of the transposed tones was ambiguous, following hypothesis B (cf. Fig. 5.2)

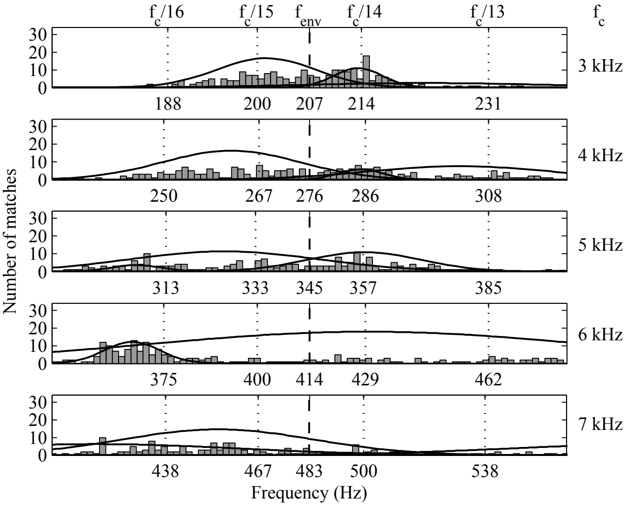


Figure 5.6: Pitch matching of the fundamental frequency of broadband pulse trains (horizontal axis) to transposed tones with a carrier frequency f_c and an envelope repetition rate f_{env}, with N=$f_c/f_{env} \approx 14.5$. See Fig. 5.4 caption for more details.

for all values of f_c. Matches were predominently close to one or several TFS-related frequencies (Fig. 5.4, dotted lines) in each individual subject. For no values of f_c were matches found to lie around the envelope repetition rate of the transposed tones f_{env} (Fig. 5.4, dashed lines). For each condition, some subjects always heard a pitch close to one single TFS-related frequency, while others focused on different TFS-related pitches in each presentation and showed multimodal distributions of matches. In the latter case, the listeners sometimes reported that the pitch of the reference stimulus suddenly changed during the matching procedure, so that they would have to make a choice between two possible matching frequencies. This confirms the overall pitch ambiguity appearing in the combined group results. When the listeners were confused by this ambiguity, they were encouraged to keep to their first impression of the stimulus pitch throughout the matching procedure.

For N=12 (Fig. 5.5), a similar pattern of results was obtained. The pitch ambiguity remained in 4 of the subjects, who showed matches related to the timing interval between the highest TFS maximum in one envelope period and the second highest TFS maximum in the following envelope period, despite the fact that, in this harmonic condition, the TFS was identical in each envelope period. However, for all subjects, the most prominent pitch was that corresponding to f_{env}.

For $N \approx 14.5$ (Fig. 5.6), the overall distributions of pitch matches showed a larger spread than for the lower values of N, indicating a less salient pitch. However, as for $N \approx 11.5$, individual pitch matches were predominently gathered around TFS-related frequencies, despite a higher number of erratic matches. For none of the conditions were the group pitch matches found to be related to the envelope period f_{env}.

5.3.2 Statistical analysis

A Gaussian mixture model was fitted to the group and individual data for each condition. The number of distributions, m, was first estimated using an iterative expectation-maximization (EM) algorithm for mixtures of 1 to 8 distributions. The EM algorithm (McLachlan and Peel, 2000, as implemented in the *gmdistribution.fit* function of the MATLAB Statistics Toolbox) was repeated 100 times for each possible value of m, and the m value returning most frequently the lowest Akaike information criterion (Akaike, 1974) was then chosen to perform 200 new iterations of the EM algorithm, each outputting the Gaussian mixture model with the maximum likelihood. For each iteration, distributions with means deviating less than 2% from one of the TFS-related frequencies were selected and a rough estimate of the proportion of TFS-related matches, P_{tfs}, was calculated as the sum of $w_i \times p_i \times (\max(0, (\mu_i - \sigma_i^2)/\mu_i))$ over all selected distributions, where p_i is the mixing proportion, μ_i the mean, and σ_i the standard deviation of distribution i. The weighting factor w_i was equal to 1 for μ_i values less than 1% away from the closest TFS-related frequency, and decreased in a $\cos^2$ fashion as μ_i deviated further than 1%, to reach 0 at a 2% deviation. The final P_{tfs} estimate for each condition was calculated as the mean of estimates obtained from all iterations, and the final fitted model parameters were extracted from the iteration for which P_{tfs} was closest to the final mean value. Parameters of the fitted models for the combined group results are given in Table 5.2, in which starred distributions indicate means deviating less than 1% from one of the TFS-related frequencies. In addition, the P_{tfs} estimates calculated from individual results are given in Table 5.3 for each condition.

In the combined group results, starred distributions were found for all conditions, except for [$N \approx 14.5$, f_c=7 kHz], and their standard deviations were overall found to increase with N and f_c. P_{tfs} also tended to become lower with increasing N and f_c, and indicated the presence of TFS-related matches in all conditions, except for [$N \approx 14.5$, f_c=6-7 kHz]. For all inharmonic conditions, none of the fitted distributions had a mean less than 1% away from the envelope repetition rate.

In the individual results, despite the lower stability of the model estimates due to the relatively small number of matches for each condition, starred distributions were found in 24 out of 35 individual data sets ($P_{tfs} > 0$ in Table 5.3). The presence of TFS-related matches also tended to disappear at high values of f_c. Distribution means close to envelope-related frequencies were found in 5 out of 30 individual data sets for the inharmonic conditions (numbers preceded by E in Table 5.3). However, in all theses cases, the standard deviations of these distributions were either too large, or the mixture proportions too low, to determine whether the corresponding matches stemmed from one envelope-related distribution, or several overlapping TFS-related distributions. Moreover, such envelope-related distribution means did not occur consistently for similar f_c values across subjects. Therefore, it is reasonable to conclude that envelope cues were never sufficient to evoke a salient pitch.

5.4 Discussion

5.4.1 TFS *vs.* envelope information

The present results demonstrate that the low pitch of transposed tones is ambiguous and corresponds to the inverse of the timing between TFS peaks close to their envelope maxima, rather than their envelope repetition rate. This is in contradiction with early assumptions according to which the peripheral auditory representation of such stimuli

Table 5.2: Best fitting Gaussian mixture model parameters to the overall pitch matching results. For each condition, the mean and standard deviation (in brackets) of each distribution are listed, as well as the proportion of the mixture represented by that distribution. A star indicates an absolute difference of less than 1% between the distribution mean and one of the TFS-related frequencies. P_{tfs} gives an estimation of the importance of TFS cues for each condition.

	N=12	N≈11.5					N≈14.5				
f_c/f_{env}	3000/250	3000/261	4000/348	5000/435	6000/522	7000/609	3000/207	4000/276	5000/345	6000/414	7000/483
1	250.5 (3.0) * 54.3%	273.4 (2.9) * 72.8%	363.8 (3.1) * 64.9%	454.2 (4.4) * 73.2%	602.4 (14.6) * 36.2%	631.0 (8.7) * 36.9%	201.2 (6.7) * 55.7%	262.0 (12.2) 54.3%	326.0 (21.2) 37.7%	429.7 (59.0) * 59.6%	454.4 (29.9) 48.5%
2	274.3 (3.7) * 24.6%	249.2 (2.6) * 24.6%	401.2 (6.3) * 18.3%	415.9 (3.2) * 15.4%	545.6 (9.7) * 31.8%	574.2 (7.2) 25.5%	213.5 (3.2) * 36.4%	285.2 (4.0) * 19.5%	357.3 (12.7) * 36.1%	366.5 (7.4) 40.4%	594.4 (47.9) 21.2%
3	248.6 (1.3) * 21.1%	293.7 (8.0) 1.8%	320.7 (11.9) 8.3%	468.0 (40.9) 11.4%	481.0 (22.6) 22.0%	704.8 (7.4) * 19.4%	222.9 (16.2) 8.5%	302.9 (16.5) * 25.3%	306.4 (5.8) 11.1%		416.2 (44.6) 20.9%
4		231.6 (0.3) * 0.4%	472.3 (12.8) 4.6%		708.3 (51.0) 10.0%	643.9 (95.1) 16.1%		396.3 (31.4) 0.9%	454.8 (30.7) * 5.1%		720.4 (17.2) 9.4%
5		263.6 (0.3) 0.4%	372.1 (3.8) 3.8%			920.5 (11.5) 2.0%			466.0 (6.5) 4.7%		
6									410.5 (5.5) 4.4%		
P_{tfs}	97	95	80	85	50	73	78	32	27	0	0

Table 5.3: Individual and group estimates of the proportion of TFS-related matches, P_{tfs}. E indicates the presence of one distribution with a mean less than 1% away from the envelope repetition rate.

Subject		1	2	3	4	5	6	All
N	f_c (kHz)							
12	3	98	99	96	98	96	94	97
11.5	3	98	87	97	99	96	85	95
11.5	4	89	87	97	99	87	22	80
11.5	5	88	95	88	96	59	89	85
11.5	6	78	14	45	82	0	59	50
11.5	7	62	18	88	67	50	13	73
14.5	3	56	80	$^{E}21$	87	48	41	78
14.5	4	30	$^{E}12$	67	0	58	0	32
14.5	5	6	42	84	27	0	$^{E}12$	27
14.5	6	64	34	0	0	0	0	0
14.5	7	0	3	1	71	$^{E}16$	$^{E}0$	0

would correspond to that of sinusoidal tones (Oxenham *et al.*, 2004). The results are in agreement with later physiological recordings showing that, for f_{env} values above 250 Hz, the temporal discharge patterns of auditory-nerve cells to transposed tones in the cat differ from that of pure tones (Dreyer and Delgutte, 2006). Moreover, the fact that the low pitch of individual transposed tones is not determined by their envelope period might explain why multiple transposed tones with a harmonic relationship between their respective envelope repetition rates do not evoke a fundamental pitch (Oxenham *et al.*, 2004). According to the present results, it is indeed very likely that the first inharmonic transposed tone used in Experiment 2 of Oxenham *et al.* (2004) had an ambiguous pitch not related to its envelope period (harmonic 3: f_c=4 kHz, N≈13.3). Moreover, the values of N for harmonic 4 (f_c=6.35 kHz, N≈15.9) and f_c for harmonic 5 (f_c=10.08 kHz, N≈20.2) may have been too high for such transposed tones to individually evoke a salient pitch. The lack of fundamental pitch found for the combination of these three stimuli may therefore be due to the lack of harmonic relationship between their individual temporal auditory representations, rather than a lack of place information.

The present results are consistent with theories of "residue" pitch as described by Schouten and de Boer (*e.g.*, de Boer, 1956b; Schouten *et al.*, 1962) and strongly suggest that their hypothesis of a low pitch corresponding to TFS peaks close to adjacent envelope maxima stays valid for tone complexes with components in high-frequency spectral regions, and for ranks of the center component up to at least 14.5, as long as a sufficient representation of TFS persists. The pitch matches obtained in the present study follow de Boer's "first approximation of the residue pitch" to a large extent (de Boer, 1956b). As the "inharmonicity index" α=(f_c mod f_{env})/f_{env} used here was slightly below 0.5, de Boer's approximation also correctly predicts the unequal amount of matches obtained under the two closest distributions on each side of f_{env} (f_c/11 and f_c/12 in Fig. 5.4, N≈11.5), with a general tendency of subjects to perceive the higher pitch, even though one cannot exclude an effect of non-sensory bias on this subjective preference. Overall, the present results are in good agreement with pitch matches earlier obtained at lower frequencies with inharmonic complex tones (de Boer, 1956b; Schouten *et al.*, 1962) and quasi-frequency-modulated signals (Ritsma and Engel, 1964). They are also consistent with the existence region of residue pitch, as measured by Ritsma (1962), if transposed tones may be treated as over-modulated signals in this context.

These results are overall consistent with the findings of Moore *et al.* (2009a), and suggest that the pitch of their stimuli was salient enough to allow discrimination in all their tested conditions, even though they did not find a drop in performance with increasing F0 at high N values, which would have been predicted by the present pitch matches. Provided their assumption of absent spectral indices holds, the present results strongly suggest that TFS information is represented and usable for complex pitch extraction in the frequency range considered in this study. However, the use of a spectral mechanism cannot be ruled out unless the inability to resolve components of the stimuli in the cochlea is verified (see below). It would remain speculative to estimate an upper frequency limit for monaural TFS representation from the results, as TFS-related matches were still obtained for [N≈11.5, f_c=7 kHz], but the calculated values of P_{tfs} suggested a reduced influence of TFS cues above 5 kHz. The low pitch salience apparent in the results for N≈14.5 contrasts with the results of Moore and Sęk (2009a), whose subjects scored above chance up to at least 8 kHz for N=14. However, a direct comparison is difficult, as the present results obtained *via* pitch matching do not necessarily imply that TFS cues cannot be used for discrimination tasks at higher f_c values. They however indicate that, for the values of N considered here, mechanisms based on envelope representations *per se*, as obtained *via*, *e.g.*, the Hilbert transformation on the input signal, do not take part in complex pitch perception. Consequently, this also suggests that envelope cues at the output of auditory filters are not involved in providing the pitch cue for discrimination of harmonic and frequency-shifted complex tones.

It remains possible that, for higher values of N, complex pitch would start corresponding to the envelope repetition rate, as found in earlier studies (Moore and Moore, 2003). However, the present results suggest that this could be explained by an increased ambiguity of TFS cues. For transposed tones, the amount of TFS peaks close to envelope maxima becomes higher as N increases, because of slower envelope fluctuations. Thus, if the most frequent ISIs in the auditory nerve represent the timing between TFS peaks close to adjacent envelope maxima, a pitch decision mechanism based on the distribution of ISIs would have to choose between a higher number of less frequent ISIs than for low values of N, where there is a low number of very frequent ISIs to choose from. In other words, as N increases, there are more ISI candidates, leading to increased pitch ambiguity. Additionally, if the absolute number of ISIs is assumed to remain constant, each candidate has a lower probability to occur, leading to less well-defined TFS-related ISIs. This increased ambiguity and degraded definition might cause the multimodal TFS-related distributions of ISIs to merge into one wide distribution centered on the envelope repetition rate of the stimulus. Therefore, pitch decision mechanisms might rely on this average estimate and return an envelope-related pitch, even though individual ISIs were determined by phase-locking to the TFS.

5.4.2 Resolvability of partials

The spectrum of the transposed tones used in this study was assumed to cover a frequency region in which all partials were unresolved. However, it is important to keep in mind that a pattern-matching mechanism to harmonic templates (Wightman, 1973; Terhardt, 1974) could give rise to similar distributions of pitch matches to the ones obtained here, if some spectral representation remained available. For instance, if auditory filters are assumed narrower than usual, the frequencies of individual components could be available to auditory pitch representations. Therefore, in order to clarify whether the lowest components of the transposed tones could be resolved by the listeners, an additional experiment was performed. Four of the above subjects were asked to compare the frequency of a pure tone to that of one of the components (the target component) of the transposed tones, using the same paradigm as Bernstein and Oxenham (2003), in which the target component was pulsed (see Appendix for detailed methods). Such a paradigm was chosen because pulsing the target component was found to lead to higher performance (Bernstein and Oxenham, 2003; Moore and Ohgushi, 1993; Roberts and Bregman, 1991; Moore *et al.*, 2006, 2009a), so that a score at chance level would give strong evidence for unresolvability. The ability to hear out the first, second, and third (center) component of all reference transposed tones with $N \approx 11.5$ used in the pitch matching experiment was measured. Results are shown in Fig. 5.7 for each value of the ratio n between target-component frequency and f_{env}, as a function of f_{env}.

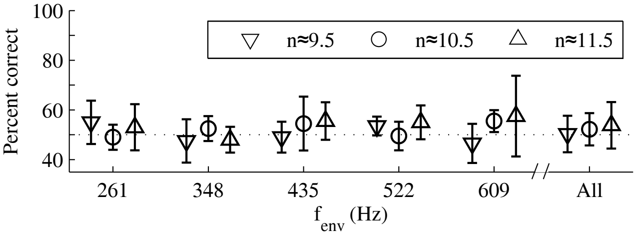


Figure 5.7: Ability of subjects to hear out the first ($n \approx 9.5$), second ($n \approx 10.5$) and third (center) component ($n \approx 11.5$) of the transposed tones. Mean results and standard deviations over all four subjects are shown for each value of n and f_{env}, and for each value of n overall. The dotted line indicates chance level.

Subjects were not found to score significantly above chance level for any value of n or f_{env} (64% correct required for significance), which suggests that individual components of the transposed tones could not be heard out. This strongly supports unresolvability of the components, all the more so as the chosen paradigm might have introduced additional cues for the listeners: They might have benefited from a release from adaptation due to the pulsing of the target component (Moore *et al.*, 2009a), and used a memorization cue because of the presence of visual feedback and the use of fixed values for f_{env}. Moreover, if a spectral representation had been responsible for the pitch matching results obtained, the scores in the partial resolvability experiment would be expected to improve as f_{env} decreased, in the same way as pitch accuracy improved with decreasing f_{env}. However, such an improvement was not found. This makes it reasonable to assume that no spectral indices were unavailable to the listeners in the pitch matching experiment.

Despite the voluntary choice of a procedure which has been argued to overestimate the maximum rank of resolvable partials (Moore *et al.*, 2009a), the use of a fixed amount of training on this partial resolvability task should be mentioned as a possible source of bias in the present study. Moreover, the fact that the training stimuli contained 9 audible components (*vs.* 5 for the transposed tones), may have made the task relatively more difficult to perform in the training runs than in the test runs. However, a repeat of the experiment in two subjects, using 5-component complex tones with $N = f_c / f_{env} = 5.5$, showed that they scored significantly above chance level for such stimuli (average scores 83.6% and 94.7%), which demonstrates the ability of these listeners to perform the task with high performance using adequately designed stimuli. Moreover, the contrast remains between the ease of the subjects in perceiving a salient low pitch and their difficulty in hearing out individual components of the exact same stimuli, even when these are pulsed on and off. Therefore, it is unlikely that sufficient information about individual component frequencies were available to retrieve the low pitch of the transposed tones.

5.4.3 Implications for pitch extraction mechanisms

Most existing models of pitch perception are based on one of two dominant suggested mechanisms: pattern matching of the retrieved frequency contents of sounds to harmonic templates stored in memory, or autocorrelation-like operations based on synchronous temporal activity in the auditory-nerve. In this section, the present results are discussed in relation to these two traditional approaches to pitch perception.

Pattern-matching models require information about the individual frequency components of a complex to be available for accurate pitch estimation. This may be achieved *via* excitation pattern cues, or *via* temporal estimates of partial frequencies.

While the latter option requires frequencies to lie below the phase-locking limit, the former does not. In both cases, if partial frequencies are retrievable, a pattern-macthing mechanism, based for instance on a Schroeder histogram (Schroeder, 1968) built from subharmonics of the stimulus partials, could in princinple correctly predict the ambiguous pitch of the inharmonic transposed tones used here. However, resolvability of the components is necessary to obtain such a histogram, and the spacing between partials of the transposed tones used in the present study always lied below 0.8 ERB. This is lower than the minimum component spacings required for resolvability of inharmonic partials, as measured by Plomp (1964) (about 1 ERB at 3 kHz and 1.9 ERB at 7 kHz) and Moore and Ohgushi (1993) (about 1.25 ERB). Moreover, the partial resolvability experiment carried out in the present study further strengthened the unresolvability assumption. Therefore, the present results do not favor the use of pattern-matching for retrieving the low pitch of transposed stimuli, if one assumes that the inability to hear out partials of a complex implies unresolvability of its components.

Time-domain models of pitch perception do neither require the use of activity patterns along the tonotopic axis, nor the presence of periodicity information for individual partials. Instead, they often base their pitch representations on autocorrelation operations, equivalent to ISI histograms built by pooling time-interval information across frequency channels. For instance, a dual-profile characterization (Ives and Patterson, 2008) of the internal representation of the transposed tones used in this study would give rise to a flat excitation pattern in their spectral profile, consistent with the unresolvability of frequency components. However, distinct maxima would appear in their temporal profile, correctly predicting the pitch ambiguity of the transposed tones. Following this approach, the ambiguous low pitch could be derived from the shortest lags giving rise to the two highest maxima in a summary-autocorrelation function (Meddis and Hewitt, 1991), and such lags would equal the most prominent TFS peak-to-peak intervals in the temporal waveform of the transposed tones. This implies that the present results are in line with such temporal models, as long as some residual phase locking to the TFS persists in the frequency range considered here.

According to this type of approach, the decrease of $\mathrm{P}_{\mathrm{tfs}}$ with f_{c} could reflect the increasing sluggishness of phase-locking with frequency. Alternatively, it could indicate a dependency on f_c of the availability of temporal lags in the autocorrelation function, should such an operation be involved. Moreover, the fact that the low pitch of a complex weakens as the number of the lowest present harmonic increases – reflected by a loss of pitch strength with increasing N – could be linked to a limitation in temporal, rather than spectral, resolution of internal auditory representations. Three main factors may thus contribute to the weakening of the low pitch with increasing N. First, the spacing between frequency components of the complex is reduced as N increases. This results in a narrower frequency range between the lowest and highest effective components of the stimulus[1], resulting in a smaller number of auditory-nerve fibers, covering a narrower range of characteristic frequencies, being activated. Thus, a mechanism pooling ISI information across the tonotopic axis would have less channels to average on. Second, individual ISIs become longer as N increases, which lowers the absolute number of ISIs in a given time window. If one assumes a temporal mechanism pooling all ISIs within a fixed time window, increasing N should thus contribute to a decrease in ISI redundancy, possibly leading to less accurate pitch representations. Finally, as mentioned above, a given ISI is likely to occur less often at higher values of N because of a larger number of TFS peaks around each envelope maximum. At low values of N, the envelope shape of transposed tones is peaky, such that there are only a couple of TFS-related ISI candidates within one envelope period, each with a high probability to occur. As N increases, the envelope shape becomes flatter, leaving space for more TFS-related ISI candidates per envelope period, each with a lower probability to occur. This increased ISI ambiguity could also contribute to the reduced pitch salience observed at the higher N value.

It might be tempting to favor an all-temporal ISI-based approach to pitch extraction, because of its independence of spectral resolvability. Auditory-nerve fiber recordings obtained in the cat (Larsen *et al.*, 2008) suggested that such ISI representations could accurately account for the F0 of concurrent complex tones, for F0s up to 900 Hz.

---

[1] Taking the background noise into account, estimates of the 3-dB excitation-pattern bandwidths for the two values of N, using the Glasberg and Moore (1990) model, were found to be of about 1.69 ERB for N=11.5 and 1.55 ERB for N=14.5. This confirms that the frequency range of excitation narrows slightly as N increases.

However, the same study showed that rate-place representations could also be valid for F0s above about 400 Hz, even in the absence of resolved components. The fact that such representations may be used for pitch discrimination does not imply that they could account for pitch estimates obtained *via* a matching procedure. Nevertheless, one cannot exclude that both types of information (time-interval and rate-place based) may contribute to pitch extraction, depending on the envelope period of the stimulus. In the present study, it remains that the low pitch was found to become less salient with increasing f_{env}, which makes the contribution of a rate-place code unlikely compared to that of an ISI representation. Furthermore, one would have expected pitch accuracy to increase with f_{env} if a spectral mechanism was involved. This is because the resolution of auditory filters increases slightly with center frequency[2], which should lead to better resolvability of partials towards high frequencies.

5.4.4 How is the fine structure represented?

The present results strongly suggest that TFS information is available to pitch extraction mechanisms at high frequencies, as a minimum in the form of accurate time-interval representations of TFS peaks around successive envelope maxima. The availability of such information to auditory neural-coding mechanisms may have two possible implications:

(1) *TFS transmission:* TFS information is transmitted from the cochlea to the auditory nerve *via* a preservation of phase locking towards higher frequency regions than previously assumed, at least for modulated signals. In such a case, a purely temporal mechanism, involving operations such as autocorrelation (Licklider, 1951; Meddis and Hewitt, 1991) or cancellation (de Cheveigné, 1998) in individual channels, would be sufficient to account for the present results, as discussed above.

(2) *TFS recovery:* Alternatively, hair-cell transduction is assumed to act as a half-wave rectifier followed by a low-pass filtering stage, after which only the envelope of the signal is processed at high-frequencies. In this case, TFS information must in

---

[2] From Equation 3 in Glasberg and Moore (1990), the quality factor of the auditory filters can be derived as $Q=F/ERB=(107.939+24.7/F)^{-1}$, which is a monotonically increasing function of F.

some way be retrieved after such a processing stage. With envelope information only, within-channel temporal mechanisms would fail to account for the present results, by always predicting a low pitch related to the envelope repetition rate. Therefore, mechanisms involving recovery of TFS cues from a comparison of internal envelope representations across frequency would be needed.

This last point raises a crucial issue: There is mathematical evidence that, for narrow-band signals, TFS and envelope cannot be entirely isolated and may be recovered from each other (Voeckler, 1966), contrary to common assumptions that they are processed independently in the auditory system. In theory, the instantaneous frequency and time of a sound signal are indeed retrievable from the pattern of time intervals between action potentials within and across fibers in the auditory nerve (Gardner and Magnasco, 2006). Such a retrieval would require neural mechanisms to act across both time and frequency on the internal spectrotemporal representation of sounds. Moreover, there is recent psychophysical (Gilbert and Lorenzi, 2006) and physiological (Heinz and Swaminathan, 2009) evidence that the envelope cues contained in a signal are partially recovered at the cochlear output, when the acoustic input only consists of the extracted TFS of the same signal. If envelope and TFS are never entirely independent and the inherent properties of cochlear filtering allow envelope recovery, it is thus likely that TFS recovery also occurs (Zeng *et al.*, 2004).

Despite a current lack of physiological evidence favoring any of these options, both remain theoretically viable. TFS transmission *via* phase-locking already occurs at low frequencies, and the idea of its persistence at high frequencies is thus easy to grasp. The existence of TFS recovery may, however, not be as straight-forward, and its importance for auditory perception deserves further investigation. Even though the present findings cannot clarify the relative validity of the TFS transmission and TFS recovery hypotheses, they demonstrate the importance of accurate TFS representation for complex pitch extraction from unresolved components, and clearly rule out the role of envelope coding as such. They further suggest that auditory prostheses ought to convey TFS information in order to restore accurate pitch perception in hearing-impaired patients.

Appendix 5: Methods for the partial resolvability experiment

A two-interval, two-alternative forced-choice procedure was used. In each trial, two 1-s intervals separated by a 375-ms silent gap were presented. The first interval contained three bursts of a 300-ms sinusoidal comparison tone with frequency f_{comp}, each including 20-ms onset and offset cosine ramps, separated by 50-ms silent gaps. The second interval contained a 1-s transposed tone, in which the target component with frequency $f_{targ}=n\times f_{env}$ was first filtered out. This target component was then gated on and off in the same way as in the first interval and, when present, had the same phase relationship to the other components as in the original transposed tone used in the pitch matching experiment. The comparison and target tones were both presented at the same level as that of the corresponding component in the original transposed tone. The same background pink noise as in the pitch matching experiment was gated on 375 ms before the first interval and off 375 ms after the second interval, with 20-ms onset and offset cosine ramps. In each trial, f_{comp} was either lower or higher than f_{targ}, with equal probability, and the absolute frequency difference between f_{comp} and f_{targ} was chosen from a uniform distribution of values between $0.035f_{targ}$ and $0.05f_{targ}$. No roving was used, *i.e.*, f_{targ} values were fixed for each condition. The task of the subjects was to identify which of the comparison and target tone was higher in frequency. The stimuli were generated and presented in the same way as in the pitch matching experiments. Fifty trials per subject were performed for each of the 15 conditions. All conditions were included within each measurement run and presented in a random order. Visual feedback was provided after each trial. As an introduction to the task, each subject participated in a training session (150 trials) in which partials at $n\approx5.5$, 7.5 or 9.5 of an inharmonic complex tone containing partials 5.5 to 13.5 were gated on and off (same intensity for all partials, overall level: 50 dB SPL).

Acknowledgments

Thanks go to Brian C. J. Moore and Hedwig E. Gockel for helpful discussions in the preparation phase of this work, Peter L. Søndergaard and Rémi Decorsière for discussions of the mathematical relationship between TFS and envelope, as well as James M. Harte for some useful comments on an earlier version of this chapter.

6

On the possibility of a spectral code for the low pitch of high-frequency complex tones

The higher harmonics of complex tones are generally considered unresolved when they interact with neighboring partials in the cochlea and cannot be heard out separately. Recent evidence suggests that the low pitch evoked by such harmonics may be coded *via* temporal fine-structure cues, even above 5 kHz. However, these conclusions rely on the assumptions that combination tones were properly masked and that the ability of listeners to hear out individual partials provides an adequate measure of resolvability. This study confirmed an influence of combination-tone audibility on the low pitch of high-frequency inharmonic complex tones containing five components with a center rank of 11.5. However, the low pitch remained salient when combination tones were masked. Despite this, there were no effects of relative component phases or dichotic presentation on the perceived pitch, providing no evidence for the use of temporal pitch cues. Moreover, similar trends were observed between pitch salience and the listeners' ability to hear out individual partials. The latter task may thus not adequately reflect resolvability in terms of complex pitch perception. Overall, the results do not rule out the use of a temporally-coded pitch, but are consistent with the use of a spectral code, which thus remains a possibility.

6.1 Introduction

Most natural sounds in our environment are complex harmonic sounds, *i.e.*, they can be decomposed in a series of frequency components that are multiples of a common fundamental frequency (F0). Such sounds usually evoke a pitch sensation corresponding to F0, even when the physical energy at F0 is removed from the signal (Seebeck, 1841). In fact, the pitch stays unchanged when additional components are removed as long as their harmonic relationship is not altered (*e.g.*, Schouten, 1940; Mathes and Miller, 1947; Davis *et al.*, 1951; Thurlow and Small, 1955). By demonstrating that the low pitch remained in the presence of low-frequency masking noise, Licklider (1954) ruled out cochlear distortion products as a valid explanation for this "missing fundamental" phenomenon. Since then, a large amount of research has been devoted to complex pitch perception, in an effort to determine the nature of the underlying auditory mechanisms (see Plack and Oxenham (2005b) for a review).

In particular, whether the spectral or temporal features of sound are used by auditory pitch mechanisms remains a central and unsolved question. The frequency analysis taking place on the basilar membrane and the tonotopic organization of the auditory pathway (Merzenich *et al.*, 1975) allow a fine internal representation of the spectral contents of sounds. Additionally, the synchronous firing of auditory-nerve cells to specific phases of the basilar-membrane vibration (Rose *et al.*, 1967) enables an accurate internal representation of the temporal features of incoming sounds. This possibility for both a high spectral and temporal resolution in the human auditory system, together with the fact that spectral and temporal information usually covary, makes it difficult to rule out the use of a specific type of information for pitch extraction.

The limitations imposed by the varying frequency-selective power of the cochlea as a function of frequency provide an important tool in the attempt to isolate spectral and temporal pitch cues. On a linear scale, the auditory filters broaden as frequency increases (Fletcher, 1940; Glasberg and Moore, 1990). This means that the low harmonics of complex sounds are resolved by the cochlea, giving rise to peaks of excitation on the tonotopic axis, while higher harmonics must interact with neighboring components within the same filter, such that their individual frequencies cannot be retrieved from the tonotopic pattern of excitation after cochlear filtering. Despite this, the latter unresolved harmonics can evoke a low pitch when presented alone (Ritsma, 1962), and this pitch is salient enough for melody recognition (Moore and Rosen, 1979). As unresolved components cannot provide place cues based on excitation, it is believed that a temporal mechanism is responsible for such a low pitch (Plack and Oxenham, 2005b).

There is general agreement that components above the tenth harmonic of a complex are unresolved, because of the inability of listeners to hear out individual partials above this rank (Plomp, 1964; Plomp and Mimpen, 1968; Moore and Ohgushi, 1993), and because of the influence of relative component phases on the perceived pitch for the higher partials (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). This means that unresolved components lie in medium to high spectral regions, where the fast temporal fluctuations, or temporal fine structure (TFS), of the output waveform of a given auditory filter are modulated in amplitude by a slowly-varying temporal envelope. In mammals, auditory-nerve cells have been shown to phase-lock to pure tones up to about 5 kHz (Rose *et al.*, 1967; Johnson, 1980), but also to the envelope of amplitude-modulated tones (Joris and Yin, 1992). Therefore, it has been questioned whether the temporal cues involved in the low pitch of unresolved components are based on TFS or envelope information. The occurence of a pitch shift when all partials of a harmonic complex tone are shifted by an equal amount on a linear frequency scale (de Boer, 1956b; Schouten *et al.*, 1962) suggests a role of TFS information, as the pitch differs even when the envelopes of the harmonic and inharmonic signals are identical. Moreover, pitch matching experiments have shown that the ambiguous pitch of inharmonic complex tones with equally-spaced components was consistent with the timing between TFS peaks close to adjacent envelope maxima in the stimulus waveform (de Boer, 1956a; Ritsma and Engel, 1964), even for unresolved components in spectral regions above 5 kHz (Santurette and Dau, 2011a).

Following such evidence, a crucial role of envelope information *per se* seems to have been ruled out. However, phase-locking in the auditory nerve of mammals is either very weak or not measurable above 5 kHz (Köppl, 1997), which makes the persistence of TFS cues implausible. Despite this, Moore and Sęk (2009a) showed that, using a pitch cue, listeners were able to discriminate harmonic and frequency-shifted bandpass-filtered complex tones only containing unresolved components above 8 kHz. The fact that they used random envelope shapes, and that providing additional envelope cues in a similar task did not increase performance (Moore *et al.*, 2009a), suggested that the use of TFS information remained possible at such high frequencies. Not only did these findings challenge the current assumptions on the frequency limit of phase-locking in humans, but they also revived the debate around the theory of "residue" pitch (*e.g.*, Schouten, 1940; de Boer, 1956a; Schouten *et al.*, 1962).

The pitch cue in Moore *et al.*'s harmonic *vs.* inharmonic discrimination task (referred to as "H/I-discrimination" in the following) has been assumed to rely on temporal mechanisms. However, this assumption remains controversial. First, Oxenham *et al.* (2009) showed that phase effects on F0 difference limens were affected by the level of the background noise used to mask cochlear distortion products. In addition to the absence of evidence for the use of TFS cues, their results indicated a potential role of combination tones in H/I-discrimination tasks. If audible, they may lie in the resolved region and their shift between the harmonic (H) and inharmonic (I) tones might provide an additional spectral cue. Moreover, the level of such combination tones and the amount of background noise required to mask them remains uncertain. Second, it is unclear to what extent the changes in excitation pattern between the H and I tones were detectable by the listeners. Moore and Sęk (2009a) filtered their stimuli in such a way that the changes did not exceed the smallest detectable change in excitation for a single component. However, changes in excitation might be easier to detect when they occur over a wide frequency range than when they are confined to a single frequency (Green *et al.*, 1987; Micheyl *et al.*, 2010a). Finally, Micheyl *et al.* (2010a) argued that differences in the temporal envelope of H and I tones at the output of individual auditory filters may also have played a role in the H/I-discrimination task.

In order to prevent the use of changes in spectral indices by the listeners, Santurette and Dau (2011a) used a pitch-matching task with inharmonic transposed tones (van de Par and Kohlrausch, 1997), to which the F0 of a broadband pulse train was matched. The pitch matches clustered around frequencies corresponding to the inverse of the timing between TFS peaks close to adjacent envelope maxima in the stimulus waveform. Moreover, the pitch of the transposed tones became less salient when the frequency or rank of the center component was raised, but the matches never corresponded to the envelope repetition rate. Together with a resolvability test showing that individual components of the stimuli could not be heard out, these results clearly suggested a role of TFS information up to carrier frequencies of 7 kHz, in line with the findings of Moore and Sęk (2009a), and ruled out a role of envelope cues as such.

Despite the consistency of such results with the TFS hypothesis, it is important to keep in mind that, had the components been resolved, a place model of pitch perception could have correctly predicted the ambiguous pitch of the transposed tones, *e.g.*, by using a histogram built from subharmonics of known partial frequencies (Schroeder, 1968; Terhardt, 1974). Before the influence of place cues can be completely ruled out in favor of TFS cues, one thus has to question whether the inability to hear out individual components of a complex implies their unresolvability. In other words, resolvability, defined as the amount of information about individual component frequencies after cochlear filtering, may have been sufficient to extract pitch, but not to hear out the individual components. After all, complex pitch extraction may be a more "automatic" process than the segregation of one component from its neighbors. Moreover, one would have to rule out audibility of combination tones for the specific stimuli used in the pitch-matching experiments to exclude their possible influence.

Given the implications of a role of TFS at high frequencies for theories of hearing, and the increasing use of H/I-discrimination tasks as a test of monaural TFS processing (Hopkins and Moore, 2007; Moore and Sęk, 2009b), the present study aimed at clarifying the possibility of a spectral code for the low pitch of high-frequency complex tones, such as those used in the aforementioned studies. More specifically, the presence of audible combination tones was evaluated, and their influence on the perceived pitch was studied directly. In order to further investigate whether the pitch could rely on place cues, the effects of relative component phases and dichotic presentation were studied, and a resolvability test was performed. Several experiments with different rationales were carried out.

In Experiment 1, an H/I-discrimination task similar to that of Moore *et al.* (2009b) was performed for F0 values between 200 and 600 Hz and harmonic ranks of the center component between 11 and 15. The general trends in performance, as a function of the frequency and rank of the center component, were compared to the pitch-matching data of Santurette and Dau (2011a).

The next two experiments were concerned with audibility of combination tones in the pitch-matching and H/I-discrimination experiments. In Experiment 2, the level of the most prominent combination tone was measured for five-component inharmonic complex tones with a center rank of 11.5 and center frequencies between 3 and 7 kHz, using a beat-cancellation method. In Experiment 3, the background pink-noise level required to mask this combination tone was estimated in a simple tone-detection task.

In order to study the effect of combination-tone audibility on the low pitch of such high-frequency complex tones, monaural pitch matches were obtained in Experiment 4 in two conditions: with sufficient background noise to mask all combination tones, and with audible combination tones. In the same experiment, the stimulus components were either in sine phase or alternating sine-cosine phase, in order to investigate the presence of phase effects on the low pitch. As changing the relative phase of the components modifies the temporal waveform but not the amplitude spectrum of the stimuli, the presence of phase effects would imply a role of temporal information and rule out the use of a rate-place code for pitch extraction.

Additionally, Experiment 5 investigated the effect of presenting the stimulus components dichotically, with every other component in the opposite ear, on the pitch matches. In such a condition, components in each ear are resolved, and the similarities or differences in pitch matches between the monaural and dichotic conditions may shed light on the spectral or temporal nature of the underlying pitch mechanisms.

Finally, the ability of listeners to hear out the individual components of the complex tones used in the previous experiments was measured in Experiment 6, using a pitch-comparison task between a pulsed sinusoid and a pulsed target component, as used by Bernstein and Oxenham (2003).

The results of the different experiments will be discussed in Chapter 7 in terms of purely spectral, purely temporal, or combined spectrotemporal theories of pitch perception.

6.2 Methods

6.2.1 Subjects

Twelve normal-hearing listeners (ages: 18-32 years) participated in the study, and only one subgroup of listeners was included in each experiment. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-KA-04149-g) and by the Institutional Review Board at the University of Minnesota. All subjects had hearing thresholds below 20 dB HL at all audiometric frequencies in both ears. In Experiments 1 to 4 and Experiment 6, they were tested monaurally in their best ear, defined as the ear with the lowest average hearing threshold between 2 and 8 kHz. All subjects had some form of musical training and played an instrument as a hobby. In the following, the different listeners are each assigned a unique number, and subject numbers are kept the same throughout the different experiments. Subject 6 was the author of the present thesis. All other listeners provided informed written consent prior to testing and were paid an hourly rate for their participation.

6.2.2 Experimental set-up

All stimuli were generated in MATLAB and presented with a 96-kHz sampling rate, either via an RME DIGI96/8 soundcard (32-bit resolution) and Sennheiser HDA200 headphones (subjects 1 to 6 and 12) or a LynxStudio L22 soundcard (24-bit resolution) and Sennheiser HD580 headphones (subjects 7 to 11), in double-walled sound-attenuating listening booths. 256-tap FIR equalization filters were applied to all stimuli, in order to flatten the frequency response of the different headphones. Additionally, hearing levels were equalized across frequency in Experiments 2 to 6. Such a correction was applied in order to compensate for the variations in absolute threshold with frequency, so that the spectral shape of the complex tones was approximately symmetric around the center component in terms of hearing level. In the following, values given in dB HL thus refer to the input sound pressure level before applying the correction.

6.3 Experiment 1: Discrimination of harmonic and inharmonic complex tones

The smallest detectable frequency shift, ΔF, between harmonic (H) and inharmonic (I) frequency-shifted bandpass-filtered complex tones, was measured in five subjects, in an H/I-discrimination task similar to that of Moore *et al.* (2009b) and Moore and Sęk (2009a). Subjects 1 to 4 had previously participated in the pitch-matching experiment described in Santurette and Dau (2011a). The present results were thus compared to discrimination thresholds predicted from the pitch matches obtained in that study.

6.3.1 Method

Stimuli

The H tones were generated by adding sinusoidal harmonics ranging from the fifth harmonic of the desired F0 to the last harmonic below 20 kHz. Random starting phases for all components produced random temporal-envelope shapes, thus limiting the availability of envelope cues for performing the task. The resulting complex was bandpass-filtered using a 512-tap FIR filter, designed after a fourth-order Butterworth response, with lower and upper cut-off frequencies at (N−2)F0 and (N+2)F0, respectively, where N is the rank of the center component of the H tone. The I tones were generated in the same fashion, except that all components were shifted upwards in frequency by ΔF. In order to minimize excitation pattern differences, the bandpass filter applied to the I tone was identical to that applied to the corresponding H tone. The overall presentation level was 50 dB SPL for all complex tones, to which a background pink noise was added in order to mask combination tones. The masking noise had a spectrum level of 13.5 dB at 1 kHz and was bandpass-filtered from 100 to 12000 Hz (512-tap FIR filter designed after a fourth-order Butterworth response).

The maximum difference between H and I excitation-patterns was calculated for the largest shift ΔF=0.5F0 with the model of Moore and Glasberg (1997), taking the effect of the background noise into account. The average difference over all conditions was 1.0 dB, and the difference never exceeded 1.8 dB, which is lower than the threshold of 3.8 dB reported in Moore and Sęk (2009a) for detecting a change in excitation when a single component is raised in level. The slightly larger excitation-pattern differences obtained in the present study, compared to those reported by Moore and Sęk (2009a), can be explained by the fact that their values were calculated for ΔF=0.18F0, whereas the present values are for ΔF=0.5F0. Moreover, the passband of the applied filter was chosen to be narrower than the one used in Moore and Sęk (2009a), such that the spectral envelope of the stimuli was similar to that of the transposed tones used in the pitch matching study of Santurette and Dau (2011a).

Procedure

A three-interval, three-alternative forced-choice (3I-3AFC) procedure was used, in which two intervals contained H tones with given F0 and N values, while the remaining, randomly-chosen interval contained a corresponding I tone. The background noise was gated on 150 ms before the onset of the first interval and gated off 150 ms after the offset of the last interval. The 450-ms tones and the noise were gated with 30-ms onset and offset cosine ramps, and the intervals were separated with 150-ms silent gaps. The listeners were instructed to report the interval which sounded different. Feedback was provided. A weighted up-down procedure (Kaernbach, 1991) was used to track the 79.4% point on the psychometric function. The threshold ΔF_{min} was tracked logarithmically, with an initial value of ΔF=0.5F0. A run was terminated after 10 reversals and the threshold value was determined from all points following the fourth reversal. During a run, if five wrong responses were given at the maximum value of the tracking variable (0.5F0), the adaptive procedure was interrupted and 55 trials were presented for ΔF fixed at 0.5F0. The percentage of correct responses over the last 50 trials was then calculated. A total of 15 conditions were measured in each measurement block: F0=[200, 300, 400, 500, 600] Hz and N=[11, 13, 15]. Each subject performed three measurement blocks, in which the presentation order of the different conditions was randomized. The final threshold was defined as the average threshold over the three blocks. The listeners were trained on at least one run for each condition, until their performance remained stable.

6.3.2 Results and discussion

In order to compare the results from the adaptive and non-adaptive procedures, d' values were calculated in both cases. For the adaptive procedure, it was assumed that d' was proportional to ΔF_{min}, as in Moore *et al.* (2009b). d' values were thus calculated by dividing the d' value tracked in the adaptive procedure (1.63) by the measured threshold, and multiplying the result by 0.5F0. For the non-adaptive procedure, percent-correct scores were converted to d' values using the table of Hacker and Ratcliff (1979).

The obtained d' values for each (F0, N) pair are plotted in Fig. 6.1 on a square-root axis. The individual data points are indicated by the corresponding subject numbers, while the gray circles and error bars represent the means and standard deviations over all subjects. Subject numbers in bold indicate that the adaptive procedure was successfully completed by that listener in the corresponding condition. For N=11, the performance of all listeners was significantly above chance (d'>0.89) for all F0 values, with the exception of subjects 2 and 5 for F0=500 Hz. For N=13, performance degraded in all listeners compared to N=11, but discrimation remained possible in at least one listener for all F0 values. For N=15, most d' values were not significantly above 0, indicating that discrimination was generally not possible for such a rank of the center harmonic.

A within-subjects ANOVA with factors F0 and N was performed on the square-root of the absolute d' values, the sign of which was restored after transformation.

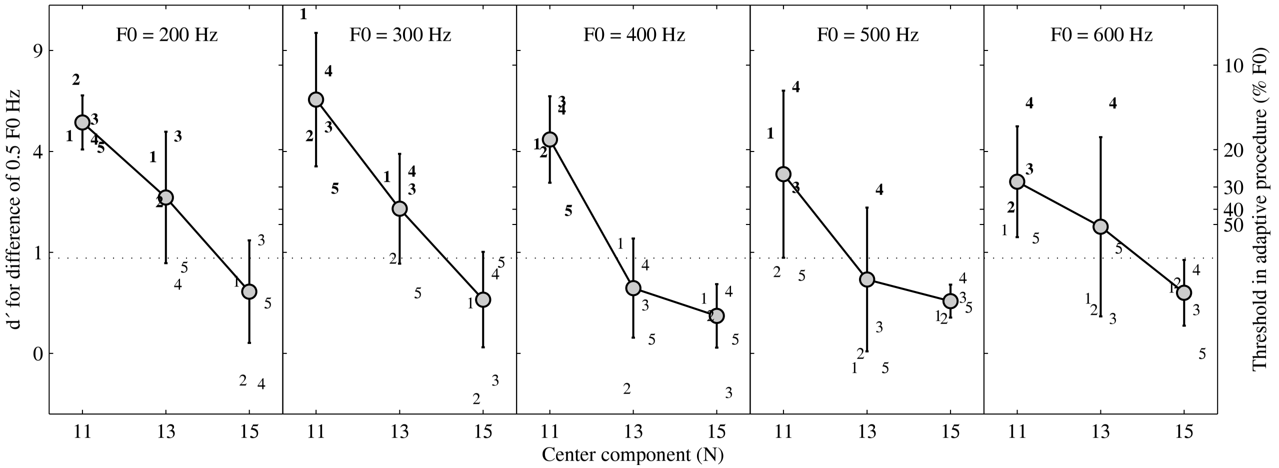


Figure 6.1: Results of the H/I-discrimination experiment, expressed as the detectability index d' for discrimination of H and I tones with ΔF=0.5F0. The right axis indicates the corresponding thresholds in the adaptive procedure. The individual results are indicated by subject numbers, as a function of N, for five F0 values, together with the mean and standard error over all subjects (gray circles and error bars). The abscissae of individual data points are staggered for readability. Bold subject numbers indicate that the adaptive procedure was completed successfully. The dotted line indicates the d' value above which performance is significantly above chance level.

A significant effect of N was found [$F(2,60)=77.61$, $p<0.0001$]. However, there was no significant effect of F0 [$F(4,60)=2.36$, $p=0.10$]. The interaction between F0 and N was significant [$F(8,60)=2.68$, $p=0.0225$]. *Post hoc* multiple pairwise comparisons confirmed that performance decreased significantly with increasing N [11-13: $p<0.0001$, 13-15: $p=0.0212$ (paired t-tests with Bonferonni correction)], while performance for any pair of F0 values did not differ significantly.

These results contrast with those of Moore *et al.* (2009b), who found that, for F0 between 50 and 400 Hz, mean d' values remained above 1 for N=15. Overall, the d' values obtained here are lower than the ones reported in their study. Two factors may explain this discrepancy. First, despite identical F0 and N values, the rank of the lowest audible component was lower in their study ($N-4$) than in the present study ($N-2$), where a narrower filter bandwidth was used. In fact, if the results are compared for equal ranks of the lowest audible component, performance in the two studies becomes similar, which might suggest that discrimination mostly relies on the lowest effective components lying within the lower skirt of the filter. Second, the background noise level used here was approximately 13 dB below the level of the most intense component in the complex tones, when expressed in dB per equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990) at 1 kHz, which is 7 dB higher than the level used by Moore *et al.* (2009b). This may also have made discrimination more difficult in the present study.

All listeners reported relying on a pitch cue to perform the task, the I tone being either higher or lower in pitch than the H tones. Two subjects mentioned that the I tone had a higher pitch more frequently than a lower pitch compared to the H tone. This is consistent with the fact that positive ΔF values were used, as these usually lead to an upward shift in pitch, which becomes more and more ambiguous as ΔF increases toward 0.5F0 (de Boer, 1956a; Patterson, 1973). One listener (subject 2) also reported that, in conditions with both high F0 and N values, an additional high-frequency pitch cue was available. Whether this could indicate audible and resolved combination tones will be investigated in Experiments 2 and 3.

6.3.3 Comparison with pitch matches

Santurette and Dau (2011a) obtained pitch matches of inharmonic transposed tones with similar spectral envelopes to those of the complex tones used here. Since the spectral region and N values of the stimuli, and the background noise and overall stimulus levels, were similar in both studies, the present H/I-discrimination results were compared to their pitch-matching data.

In order to express the results of the two experiments in similar terms, discrimination thresholds were predicted from the pitch matches using an approach based on estimation theory (Edgeworth, 1908) and described in Micheyl *et al.* (2010b). For transposed tones, a shift of a given amount ΔF in the carrier frequency $\mathrm{f_c}$, which equals the center frequency of the complex, is equivalent to shifting all frequency components by ΔF. For each condition, the threshold $\Delta\mathrm{F_{min}}$ for detecting such a shift was thus estimated as

$$\Delta\mathrm{F_{min}} = \frac{d'_{\mathrm{tracked}}}{\sqrt{I_{\mathrm{f_c}}}},$$

where d'_{tracked}=1.63 is the d' value tracked in the H/I-discrimination experiment and $I_{\mathrm{f_c}}$ is the Fisher information of the $\mathrm{f_c}$ parameter. $I_{\mathrm{f_c}}$ was evaluated numerically as

$$I_{\mathrm{f_c}} = E\left[\left(\frac{\partial}{\partial \mathrm{f_c}} \ln pdf(X,\mathrm{f_c})\right)^2\right],$$

where $pdf(X,\mathrm{f_c})$ is the probability density function derived from the parameters of the Gaussian mixture models fitted to the pitch matches X of Santurette and Dau (2011a). Distributions for which the mixing proportion μ corresponded to a single data point (μ=0.4%) were ignored in the calculation of the pdf. In order to estimate the partial derivative in the above equation, it was assumed that, for small shifts ΔF, the means of the distributions of pitch matches were multiplied by $1+\Delta\mathrm{F}/\mathrm{f_c}$, while their standard deviations and mixing proportions remained constant. Numerical differentiation was then applied, using a three-point estimation of the slope of ln $pdf(X,\mathrm{f_c})$, calculated as

$$\frac{\ln pdf(X,\mathrm{f_c}+\Delta\mathrm{F}) - \ln pdf(X,\mathrm{f_c}-\Delta\mathrm{F})}{2\Delta\mathrm{F}},$$

with ΔF=0.01 Hz. The resulting ΔF_{min} values were then converted to d' values in the same fashion as the thresholds obtained in the H/I-discrimination experiment: $d'=d'_{\mathrm{tracked}}/\Delta F_{\mathrm{min}} \times 0.5 f_{\mathrm{env}}$, where f_{env} is the envelope repetition rate of the transposed tone of interest.

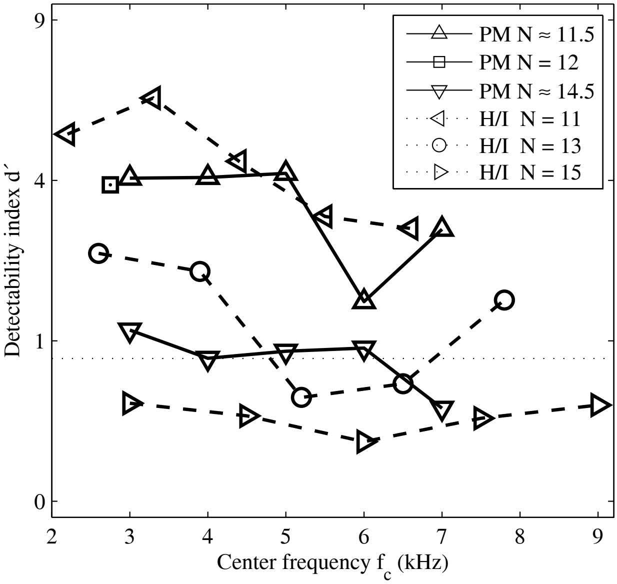


Figure 6.2: Comparison of d' values predicted from pitch matches (PM, solid lines) to those obtained in the H/I experiment (H/I, dashed lines), for different values of N, as a function of the center frequency of the complex tones f_c. The dotted line indicates the d' value above which performance is significantly above chance level. The data point for N=12 and f_c=3 kHz is slightly staggered for readability.

These d' values predicted from pitch matches are provided in Fig. 6.2 (PM, solid lines) as a function of f_c, together with the mean H/I discrimination results (H/I, dashed lines), for the different values of N used in both experiments. It appears that the predicted performance from the pitch matching data is broadly consistent with the H/I-discrimination results. This follows the subjects' reports that pitch was the main cue in the H/I-discrimination experiment. The predicted d' values suggest that the pitch salience of the stimuli was sufficient to serve as a cue for discrimination for f_c values up to at least 6 kHz as long as N remained below about 14.5. Above this higher limit for N, H/I-discrimination is not predicted to be significantly above chance for any f_c value. The present H/I-discrimination data show that human performance is not always as good as these predictions, as discrimination already drops below the significance level for N=13 and f_c>5 kHz. This may partly be due to the choice of a 3-AFC procedure in the present study, in which the pitch ambiguity of the I tone may have made human discrimination more difficult than in a change detection task, such as that used by Moore and Sęk (2009a). Poorer performance may also have arised from the use of random component phases in the H/I-discrimination experiment, *vs.* fixed component phases in the pitch matching experiment. It is also worth noting that this overestimation of d', as predicted from the pitch matching data, only arises for $f_c \geq 5$ kHz, where performance does not decrease consistently with increasing N as for $f_c < 5$ kHz.

Despite the general trend of a poorer performance with increasing f_c, reflected in both the pitch matching and the discrimination data, it can be seen that d' values increase again for f_c>6 kHz in some conditions (PM: N≈11.5, H/I: N=[13, 15]). This was also reflected in the P_{tfs} estimate of the proportion of TFS-related pitch matches calculated by Santurette and Dau (2011a), which was higher for f_c=7 kHz than for f_c=6 kHz (N≈11.5). These P_{tfs} values and the d' predictions shown above ought to be considered as rough estimates allowing only limited quantitative interpretation. However, the fact that the H/I-discrimination results showed a similar behavior suggests that this increase in performance at high f_c values is not a negligible effect accountable for by variability in the data, and deserves further investigation. As mentioned above, the fact that one listener reported the presence of an additional cue for H/I-discrimination in the high-frequency conditions suggests that this could be due to the presence of audible, resolved combination tones. This is also suggested by the fitted distributions to the pitch matches of Santurette and Dau (2011a), the means of which are slightly shifted at f_c=7 kHz compared to lower f_c values (N≈11.5). Confirming or ruling out the presence of combination tones was thus the aim of Experiments 2 and 3 described below.

6.4 Experiment 2: Level of the most prominent combination tone

In their study, Santurette and Dau (2011a) assumed that the background noise level was sufficient to mask combination tones, based on the available literature about behavioral measures of the $2f_1-f_2$ combination-tone level (*e.g.*, Goldstein, 1967; Hall, 1972; Zwicker, 1981). However, such levels were measured with two-tone complexes. For the five-component complexes of interest here, a simple estimation[1] based on the data of Goldstein (1967) suggests that the level of the combination tone lying at $f_c \times (N-3)/N$ may have been nearly as high as that of the lowest physically-present component. Therefore, this experiment aimed at estimating the level of the most prominent aural cubic difference tone below the physical spectrum of 5-component complex tones such as those used in Experiment 1 and in the pitch matching experiment of Santurette and Dau (2011a). Five listeners participated in this part of the study.

6.4.1 Method

Stimuli

The inharmonic complex tones consisted of five primary tones and had a center frequency f_c=[3, 5, 7] kHz. The ratio N between f_c and the envelope repetition rate (or component spacing) f_{env} was always equal to 11.5. In all conditions, the level of the center component was 46.6 dB HL, that of the components at $f_c \pm f_{env}$ was 44.0 dB HL, and that of the components at $f_c \pm 2f_{env}$ was 32.8 dB HL, leading to an overall stimulus level of 50.0 dB HL. Such levels were chosen for comparison purposes, as they were similar to the component levels of the transposed tones used in Santurette and Dau

---

[1] According to the data of Goldstein (1967), the levels of the $2f_1-f_2$, $3f_1-2f_2$, and $4f_1-3f_2$ combination tones are, respectively, approximately 15, 30, and 40 dB below those of two primary tones of equal level with a frequency ratio f_2/f_1=1.1. For the stimuli described in section 6.4.1, a conservative calculation in which all possible combination tones would have an additive effect gives a value of 32.4 dB HL for the estimated maximum level of the combination tone at $f_c \times (N-3)/N$.

(2011a). In the present study, however, the components were generated independently in order to control their relative phases. An example of the temporal waveform and frequency spectrum of the stimuli for the f_c=5 kHz conditions is given in Fig. 6.3. The components were added either in sine phase (SIN configuration: 0 starting phase for all components, left panel in Fig. 6.3), in alternating phase (ALT configuration: $\pi/2$ starting phase for components at $f_c \pm f_{env}$, 0 starting phase for other components, right panel in Fig. 6.3), or in the same phase relationship as that of the transposed stimuli mentioned above (TS configuration).

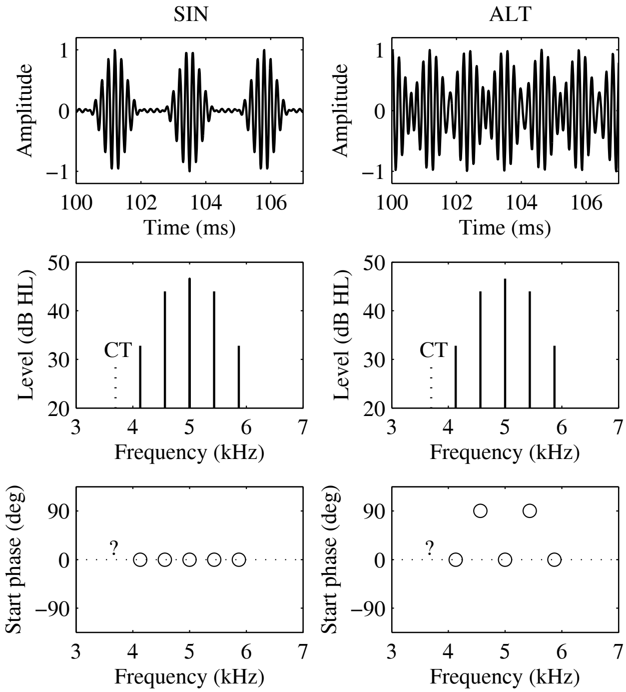


Figure 6.3: Temporal waveform (upper pane), component levels (medium pane) and component starting phases (lower pane) for two complex tones with center frequency f_c=5 kHz and component spacing $f_{env}=f_c/N$, with N=11.5. Components are either added in sine phase (SIN configuration, left column) or in alternating phase (ALT configuration, right column). CT indicates the frequency of the most prominent combination tone, $f_{CT}=f_c\times(N-3)/N=f_c-3f_{env}$.

Procedure

The level and phase of a pure tone cancelling the combination tone (CT) with frequency $f_{CT}=f_c\times(N-3)/N$ (see Fig. 6.3) was estimated using a beat-cancellation method (Schouten, 1938; Goldstein, 1967). The level and phase of a cancellation tone at f_{CT} were adjusted such that the clear beat created by a probe tone at $f_{CT}-4$ Hz became inaudible. Each experimental run was divided into two parts. In part one, no cancellation tone was added to the primary tones of the complex, and the listeners were instructed to adjust the level of the probe tone such that the beating cue was maximal. In part two, the level of the probe tone was fixed to that determined in part one (L_{max}), and the listeners were instructed to adjust the level and phase of the cancellation tone until the beat was totally eliminated. In both parts, levels could be varied between 0 dB HL and 50 dB HL, in steps of 5 dB, 2 dB, or 0.5 dB. The starting level was 10 dB HL for the probe tone and $L_{\text{max}}-5$ dB for the cancellation tone. In part two, the phase could be varied in steps of $\pi/2$, $\pi/8$, or $\pi/32$, and the starting phase was random. In each presentation, the probe tone was gated on 500 ms after stimulus onset, for a total stimulus duration of 1.5 s. All tones were gated with 30-ms onset and offset cosine ramps. Each experimental block contained one run for each condition, which were presented in a random order. Each subject performed a minimum of 3 blocks. The results of pilot tests suggested that the ALT configuration gave rise to either a very weak beating cue or consistently lower cancellation levels than the SIN and TS configurations, which gave rise to similar outcomes. Moreover, Oxenham *et al.* (2009) also found lower CT levels for 3-component complexes in the ALT configuration than in the SIN configuration. Therefore, the SIN configuration was chosen to perform the test in the whole subject group, and only subject 6 additionally collected a whole data set for the ALT and TS configurations.

6.4.2 Results and discussion

The individual and average results are presented in Table 6.1, where the obtained L_{max} values are given, together with the cancellation levels L_{cancel} and cancellation phases

P_{cancel} required to make the beating cue disappear. With this method, it is assumed that L_{cancel} reflects the approximate level of the CT.

Table 6.1: Results of the combination tone measurement. For each condition, the level of the probe tone for a maximal beating cue L_{max} is given in dB HL, as well as the level L_{cancel} in dB HL and phase P_{cancel} in degrees of the cancellation tone required to eliminate the beating cue. The mean and standard deviation (in brackets) are given for each subject, as well as the mean and standard deviation over all subjects for the SIN configuration.

Subject	Conf.	$\mathrm{f_c}$ (kHz)	L_{max}	L_{cancel}	P_{cancel}
6	ALT	3	22.7 (1.5)	20.8 (1.0)	106.8 (13.7)
		5	29.0 (1.0)	20.7 (0.8)	16.9 (5.2)
		7	27.7 (1.5)	18.7 (2.8)	101.6 (5.0)
6	TS	3	28.3 (1.7)	30.7 (0.8)	59.5 (5.1)
		5	32.0 (1.2)	26.8 (1.3)	111.7 (2.2)
		7	34.7 (1.7)	29.3 (3.2)	18.1 (6.9)
6	SIN	3	27.7 (1.2)	30.8 (1.0)	131.7 (8.8)
		5	33.3 (2.7)	29.5 (2.2)	311.0 (8.2)
		7	36.7 (1.7)	31.3 (3.0)	318.4 (4.7)
2	SIN	3	30.0 (2.0)	23.7 (3.2)	127.3 (38.0)
		5	34.7 (4.7)	26.7 (3.1)	277.1 (5.1)
		7	41.3 (2.3)	26.0 (4.6)	106.8 (7.0)
3	SIN	3	31.0 (2.6)	34.0 (1.0)	340.8 (78.3)
		5	24.7 (2.3)	30.5 (4.5)	40.0 (50.7)
		7	25.8 (1.8)	9.7 (5.8)	39.8 (32.3)
7	SIN	3	34.7 (3.4)	33.5 (2.2)	129.6 (6.0)
		5	30.2 (1.0)	29.2 (0.3)	252.1 (4.3)
		7	20.3 (1.9)	18.2 (1.0)	341.8 (14.3)
8	SIN	3	20.8 (1.0)	20.0 (3.6)	305.2 (29.6)
		5	19.9 (3.0)	27.8 (7.7)	334.0 (61.9)
		7	22.4 (2.0)	21.4 (3.6)	3.5 (46.8)
All	SIN	3	28.8 (5.1)	28.4 (6.3)	206.9 (106.7)
		5	28.6 (6.2)	28.7 (1.5)	314.8 (57.0)
		7	29.3 (9.2)	21.3 (8.2)	18.1 (58.0)

For the SIN configuration, L_{cancel} was on average found to be 4.4, 4.1, and 11.5 dB below the level of the lowest component of the physical stimulus, for center frequencies of 3, 5, and 7 kHz, respectively. A large variability was observed across subjects, indicating that CT levels are highly subject-dependent, as observed by Zwicker (1981). A repeated-measures ANOVA on the last three data points obtained for each (subject, $\mathrm{f_c}$) pair confirmed a significant effect of subject [$F(2,30)=18.06$, $p<0.0001$]. The effect of $\mathrm{f_c}$ [$F(4,30)=5.37$, $p=0.0022$] and the interaction between subject and $\mathrm{f_c}$ [$F(8,30)=8.43$, $p<0.0001$] were also significant.

For similar ratios between adjacent component frequencies to the ones used here (≈1.1), data obtained with two-tone complexes at lower frequencies suggest that the level of the $2f_1-f_2$ CT lies about 10 to 20 dB below that of equal-level primaries (Goldstein, 1967; Hall, 1972), and that this level remains similar at high frequencies (Zwicker, 1981). Comparatively higher CT levels were obtained here in many of the conditions, indicating an additive effect of the different combination tones generated by the five-component complexes. In some cases, the measured levels were in fact similar to the maximum CT level estimated above. In some subjects and conditions, the level of the most prominent CT was thus as high as that of the lowest stimulus component.

A two-way ANOVA on the results of subject 6 confirmed a sigificant influence of the configuration of relative component phases on the CT level [$F(2,18)=70.64$, $p<0.0001$], while there was no effect of f_c [$F(2,18)=1.75$, $p=0.20$] and no interaction between configuration and f_c [$F(4,18)=1.38$, $p=0.28$]. *Post hoc* t-tests confirmed that the CT level was significantly lower in the ALT configuration than in the TS and SIN configurations [$p<0.0001$], but similar in the TS and SIN configurations [$p=0.15$].

Over all conditions, the difference between L_{max} and L_{cancel} was not significant [$p=0.0553$] (paired t-test)], suggesting that L_{max} may also serve as an estimation of the CT level. Only for $f_c=7$ kHz was L_{max} always larger than L_{cancel}. One explanation for this could be that the listeners were biased by the overall loudness of the stimulus when judging the strength of the beating in part one of the experiment, such that they tended to perceive the beats of louder stimuli as stronger.

Finally, even though L_{cancel} was the outcome measure of interest in the present study, one may note that the cancellation phase P_{cancel} also showed a dependence on subject, f_c, and stimulus configuration, as reflected in the values reported in Table 6.1.

6.5 Experiment 3: Noise level required to mask combination tones

In order to evaluate whether all combination tones were properly masked in Experiment 1, the amount of background pink-noise required to mask the most prominent cubic combination tone was measured in a masking experiment performed by six listeners.

6.5.1 Method

The background-noise level required to mask a pure tone with the same frequency and level as that of the most prominent CT was measured, using a 3I-3AFC procedure. Pure tone frequencies of $f_c \times (N-3)/N$ were used, with f_c=[3, 5, 7] kHz and N=11.5. The 700-ms intervals all contained noise, while the target 300-ms tone was temporally centered in one randomly chosen interval. The noises and the tone were gated with 30-ms onset and offset cosine ramps, and the intervals were separated by 300-ms silent gaps. The level of the pure tone was fixed at 32 dB HL, *i.e.*, above the average CT level estimated in Experiment 2. The 1-kHz spectrum level of a pink-noise masker, generated as in Experiment 1, was varied adaptively with a weighted up-down method (Kaernbach, 1991) tracking the 50% point on the psychometric function. The initial spectrum level of the noise was 0 dB HL. Stepsizes of 5 dB, 2 dB, and 0.5 dB were used and the value decreased after each lower reversal. A run was terminated after 10 reversals and the threshold value was determined from all points following the fourth reversal. Runs in which the standard deviation of the last 6 reversal points exceeded 1 dB were discarded and repeated. Each subject performed a total of 3 runs per condition and the final threshold was defined as the average threshold over all runs. The conditions were presented in a random order. Feedback was provided.

6.5.2 Results and discussion

The obtained results are given in Table 6.2. For f_c=[3, 5] kHz, all listeners obtained thresholds below the spectrum level of 13.5 dB for the background pink noise used in Experiment 1 and in Santurette and Dau (2011a). This indicates that such a level could reasonably be assumed sufficient to mask combination tones in conditions where f_c was 5 kHz or below. However, half of the listeners obtained thresholds above 13.5 dB for f_c=7 kHz, suggesting that combinations tones may not have been masked in all listeners for f_c values above 5 kHz. Although the results of Experiment 2 showed that the CT level was on average lower for f_c=7 kHz than for lower f_c values, this was not the case in all subjects. Therefore, the presence of audible combination tones cannot be ruled out for $f_c>5$ kHz.

Table 6.2: Pink-noise spectrum levels at 1 kHz required to mask a pure tone at 32 dB HL with frequency $f_c \times (N-3)/N$, for N=11.5. The individual thresholds as well as the mean and standard deviation (SD) over all subjects are given.

Subject	f_c=3 kHz	f_c=5 kHz	f_c=7 kHz
6	12.7	11.1	16.3
7	10.5	12.3	14.6
8	11.1	9.3	11.8
9	11.1	9.3	15.0
10	9.9	9.0	12.6
11	10.8	8.1	12.9
Mean	11.0	9.8	13.8
SD	1.0	1.5	1.7

These results are consistent with a general increase in pitch salience and in performance of the listeners in H/I-discrimination above 6 kHz. Because of the dependence of CT level on relative component phases (*cf.* Experiment 2), they provide an additional explanation for the overestimation of performance in the H/I-discrimination task above 5 kHz, when predicted from the pitch matching data (*cf.* Fig. 6.2). This is because of the use of the TS configuration in the pitch matching experiment *vs.* random component phases in Experiment 1, in which combination tones could thus be expected to have a lower level on average.

6.6 Experiment 4: Influence of combination-tone audibility and relative component phases on pitch matches

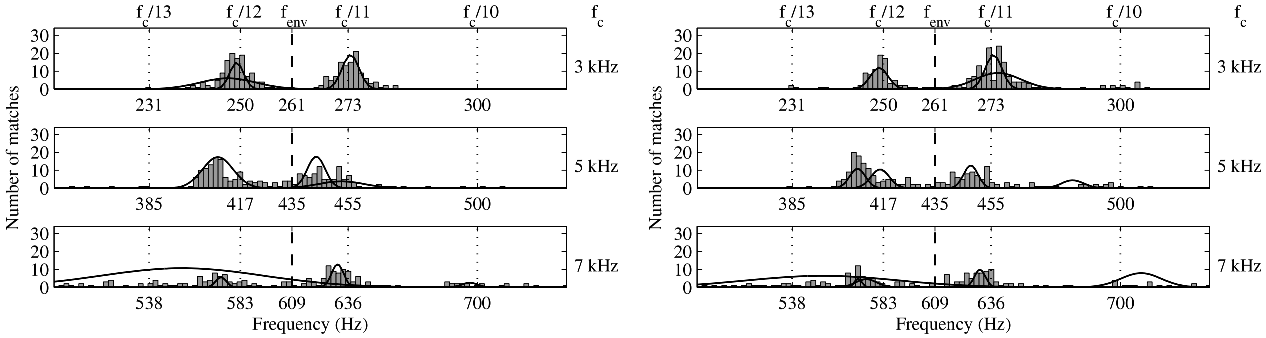
This experiment directly investigated the influence of CT audibility and relative component-phase configuration on the low pitch of high-frequency complex tones, for f_c=[3, 5, 7] kHz and N=11.5. A pitch matching experiment similar to that of Santurette and Dau (2011a) was carried out, in which the reference stimuli were 5-component complex tones in the SIN or ALT configurations illustrated in Fig. 6.3. Shackleton and Carlyon (1994) showed that the pitch of complex tones in the ALT configuration differed from that of complexes in the SIN configuration when the lowest harmonic had an approximate rank of 16, but not when this rank was lowered to about 6, for F0=250 Hz. According to their definition of resolvability, the present stimuli would lie around the upper limit of the transition region between resolved and unresolved components, suggesting that phase effects should occur. This experiment investigated the existence of such effects for the 5-component complex tones of interest here. The effect of CT audibility was investigated by comparing pitch matches in conditions where the background noise level was sufficient to mask all CT to conditions in which no background noise was present in spectral regions containing the most prominent cubic difference tones. Six listeners (subjects 2, 3, 6, 7, 9, and 12) participated in the experiment.

6.6.1 Method

The reference stimuli were identical to those used in Experiment 2 and generated in the same way as described in section 6.4.1. The listeners were asked to adjust the fundamental frequency f_p of broadband pulse trains, which were generated by adding harmonic cosine tones of f_p, starting at the fifth harmonic, then bandpass filtered between 2 and 10 kHz using a 512-tap FIR filter designed after a fourth-order Butterworth response. f_p could be varied in steps of 4 semitones, 1 semitone, or 1/4 semitone, and the starting value for each presentation was randomly chosen from a uniform distribution of values between 0.8f_{env} and 1.2f_{env}. The subjects were able to play the 500-ms reference and matching stimuli as many times as they wished, with no lower or upper limit for f_p, until they were satisfied with the match. All stimuli were gated with 30-ms onset and offset cosine ramps, and the overall level of the pulse trains was 55 dB HL. A background pink noise, generated as in Experiment 1, was played continuously throughout the matching procedure. In the "masked CT" conditions, the 1-kHz spectrum level of the noise was set to 13.5 dB HL for f_c=[3, 5] kHz and to 17.0 dB HL for f_c=7 kHz, which should be sufficient to mask all cubic difference tones in all subjects, according to the results of Experiment 3. In the "audible CT" conditions, the upper cut-off frequency of the noise was lowered to 700 Hz, such that the most prominent difference tone at f_{env} remained masked, while other combination tones were audible. Each listener performed 10 runs of 60 matches each, *i.e.*, 10 matches per (f_c, configuration) pair, presented in a random order. The CT were masked in half of the runs and audible in the other half, with alternating "masked CT" and "audible CT" runs. Matches from the last 4 runs were included in the final results for each subject. Before the experiment, it was ensured that the pitch accuracy of novice subjects was similar to that of those already familiar with the task, by collecting pure-tone matches to reference broadband pulse-trains in a preliminary experiment, as described in Santurette and Dau (2011a).

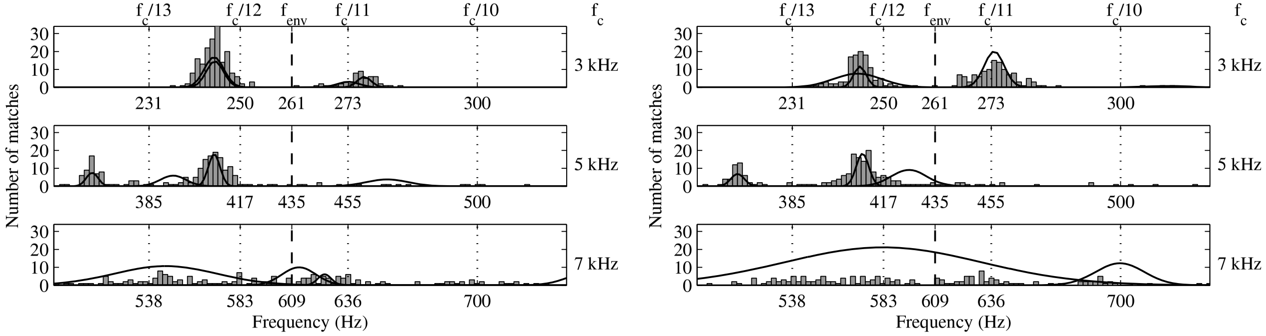
6.6.2 Results and discussion

The distributions of matches for the whole subject group (240 matches per condition) are illustrated in Fig. 6.4, using histograms with a bin width of f_{env}/250. For each condition, Gaussian mixture models were fitted to the data using the same procedure as Santurette and Dau (2011a). The model outputs are represented by solid lines in Fig. 6.4 and the corresponding parameters are detailed in Table 6.3. In order to relate the results to those of Experiment 1, d' values were predicted from the pitch matches in the same way as described in section 6.3.3. These values are given in Table 6.3.



(a) SIN configuration, masked CT.

(b) ALT configuration, masked CT.



(c) SIN configuration, audible CT.

(d) ALT configuration, audible CT.

Figure 6.4: Pitch matching of the fundamental frequency of broadband pulse trains (horizontal axis) to five-component high-frequency complex tones with a center component at f_c and an envelope repetition rate f_{env}=f_c/11.5, in four conditions: for masked (higher panels) or audible (lower panels) combination tones, and for component phases in the SIN (left panels) and ALT (right panels) configurations. The total distribution of pitch matches for all six subjects is shown (histograms and fitted Gaussian mixture models), with 40 matches per condition per subject. The height of the fitted distributions reflects their mixing proportions. The vertical dashed lines indicate f_{env} for each condition, while the dotted lines indicate subharmonics of f_c.

Table 6.3: Best fitting Gaussian mixture model parameters to the overall pitch matching results of Experiment 4. For each condition, the mean and standard deviation (in brackets) of each distribution are listed, as well as the proportion of the mixture represented by that distribution. The predicted d' values for a discrimination task are also listed at the bottom of each column. Stars indicate d' values significantly above chance level for a task similar to that of Experiment 1.

	Masked CT						Audible CT					
f_c/Conf.	3 kHz/SIN	3 kHz/ALT	5 kHz/SIN	5 kHz/ALT	7 kHz/SIN	7 kHz/ALT	3 kHz/SIN	3 kHz/ALT	5 kHz/SIN	5 kHz/ALT	7 kHz/SIN	7 kHz/ALT
1	247.4 (6.0) 15.1%	249.1 (1.9) 30.0%	408.7 (5.4) 43.3%	407.6 (2.8) 26.9%	554.7 (40.2) 26.7%	553.5 (37.0) 16.1%	244.5 (2.1) 41.1%	244.5 (5.1) 19.1%	364.6 (2.1) 18.6%	365.4 (2.4) 17.1%	546.6 (25.1) 26.8%	583.1 (48.5) 52.9%
2	249.2 (2.1) 37.0%	273.3 (3.1) 47.4%	443.3 (9.9) 43.8%	415.5 (10.5) 25.9%	573.8 (6.9) 13.9%	570.8 (3.5) 12.0%	244.7 (3.3) 36.0%	245.0 (1.7) 29.2%	393.1 (20.6) 14.9%	409.3 (4.9) 45.1%	612.3 (69.1) 25.0%	699.7 (143.2) 30.7%
3	273.3 (3.2) 47.5%	274.2 (23.2) 22.6%	452.2 (65.9) 9.5%	447.4 (6.3) 31.5%	631.2 (8.8) 31.9%	574.9 (26.7) 11.3%	272.8 (5.3) 7.8%	273.4 (4.4) 49.6%	407.5 (4.2) 43.7%	425.6 (38.4) 22.8%	624.8 (11.1) 15.2%	2105.4 (589.6) 16.5%
4	359.8 (0.0) 0.4%		892.4 (38.2) 3.3%	483.2 (15.2) 11.1%	696.3 (14.8) 6.5%	630.8 (7.3) 24.6%	276.2 (2.1) 14.8%	309.9 (20.0) 2.0%	468.4 (53.9) 9.5%	836.2 (18.6) 2.1%	755.9 (119.2) 18.5%	
5				885.9 (128.6) 4.6%	771.7 (80.4) 10.2%	710.2 (101.0) 19.8%	328.9 (0.0) 0.4%		769.9 (0.0) 0.4%	941.6 (0.0) 0.4%	1907.9 (319.0) 14.6%	
6					1195.6 (91.1) 7.6%	1218.4 (138.3) 5.0%			824.9 (20.1) 3.3%	1238.9 (0.0) 0.4%		
7					1779.8 (152.1) 3.3%	1773.3 (147.8) 5.3%			1656.0 (28.3) 9.6%	1655.0 (24.2) 9.6%		
8						2040.6 (295.4) 6.1%				1876.2 (57.2) 2.5%		
d'	3.7*	3.7*	2.3*	3.1*	2.0*	2.6*	4.0*	3.4*	4.1*	3.6*	0.8	0.4

Effect of phase configuration

As can be observed by comparing the left and right panels in Fig. 6.4, there was no dramatic effect of phase configuration on the distribution of pitch matches in any of the conditions. If the pitch had relied on the timing between the most prominent TFS peaks in the stimulus waveform, one would have expected the distribution means for the ALT configuration to have approximately double values compared to those for the SIN configuration (*cf.* Fig. 6.3). However, this was clearly not the case, as reflected by the distribution means listed in Table 6.3. Only few matches lay approximately one octave higher than f_{env}. Moreover, when present, such matches occured for both the SIN and ALT configurations and represented only a small proportion of the data (*cf.* Table 6.3: Masked CT/3-5 kHz, Audible CT/5 kHz). In fact, an analysis of the individual data showed that these "octave" matches were almost all obtained in the same subject, and that they were totally absent in four of the listeners. This indicates that they probably were the result of octave confusions by two of the listeners, which may have arised since there was no upper limit for the pulse-train F0 in the matching procedure.

In order to statistically compare the obtained distributions of pitch matches, two-sample Kolmogorov-Smirnov tests were performed on the individual data sets for each condition. Out of 36 comparisons (6 conditions for each of the 6 subjects), 25 showed no significant difference between the SIN and ALT configurations ($p>0.05$), while 4 showed a borderline-significant difference ($p=0.0431$), and only 7 a significant difference ($p<0.01$). Moreover, in all 7 data sets for which a significant difference was found, this was never the result of a difference in the means of the distributions of pitch matches. Instead, it was either due to a change in mixing proportions between several ambiguous pitches, in which case a higher pitch was generally preferred in the ALT configuration, or to a lack of salient pitch.

In summary, the configuration of relative component phases had no effect on the perceived pitch of the complex tones considered here. Moreover, this was the case for both masked and audible combination tones. Such results are consistent with those of Houtsma and Smurzynski (1990), who found that complex tones in Schroeder phase gave rise to poorer melody identification and F0 discrimination than sine phase complexes when the rank of the lowest harmonic was 13, but to similar performance when this rank was 10. Despite some uncertainty about CT audibility in their study, this may not have been a crucial factor, as the present results confirm the absence of phase effects for a lowest rank of 9.5, even when all CT are properly masked. If one requires phase effects on pitch when defining unresolvability, the lowest components of the present stimuli were thus resolved.

Had phase effects been present here, this would have indicated the use of temporal cues for pitch perception. However, the absence of such effects does not rule out the possibility of a temporal mechanism. The implications of the lack of phase effects on the temporal or spectral nature of pitch mechanisms will be discussed in Chapter 7.

Effect of CT audibility

A comparison of the upper and lower panels in Fig. 6.4 reveals an influence of CT audibility on the perceived pitch. Several effects were found.

First, the distribution means generally lay further away from subharmonics of f_c for audible-CT than for masked-CT conditions. This was mainly observed for distributions around $f_c/12$, for which a clear downward pitch shift between masked and audible CT was observed in four subjects (2, 7, 9, 12) for f_c=[3, 5] kHz, and for subject 6 at 3 kHz. Such pitch shifts are in line with those reported by Smoorenburg (1970), who explained these shifts by the fact that the center of gravity of the internal spectral representations is shifted downwards on a tonotopic axis by audible combination tones. According to his data, based on two-component complexes at 40 dB SL, one could expect the lowest audible CT in the present study to lie at $5.5\times f_{env}$, and the “effective frequency” for pitch detection to be about $7\times f_{env}$, leading to predicted pitches of 242 and 404 Hz for respective f_c values of 3 and 5 kHz. Such values are close to those obtained here by the above subjects. However, pitch shifts between masked and audible-CT conditions were not always observed in the present study. For distributions around $f_c/11$, there was generally no pitch shift, except a slight upward shift in subject 3 for f_c=3 kHz. In addition, subjects 6 and 7 also showed distribution means lower than $f_c/12$ in the masked-CT condition for f_c=5 kHz, which is reflected in the group data (Fig. 6.4, upper left panel). According to the results obtained by these two subjects in Experiments 2 and 3, it is unlikely that this was due to an insufficient level of the masking noise in this condition. Moreover, such shifts occured for both the SIN and ALT configurations, despite a lower CT level for the ALT configuration, further suggesting that they were not due to CT audibility.

Second, the mixing proportions of distributions with means below f_{env} were always higher for audible CT than for masked CT, for f_c=[3, 5] kHz. This indicates that, in the presence of several ambiguous pitches, the listeners tended to choose a lower pitch when combination tones were audible. Such a trend was clearly visible in the individual results of subjects 2, 6, 7, and 12. Whether this could be predicted by the use of specific pitch mechanisms will be discussed in Chapter 7. As the presence of combination tones considerably extends the aural spectrum toward lower spectral regions, it is also possible that a difference in timbre between the masked-CT and audible-CT conditions played a role in the observed change in pitch preference.

Third, for f_c=7 kHz, the pitch was less salient when the background noise was absent (audible CT) than when it was present (masked CT). This somewhat counterintuitive result is clearly reflected by the predicted d' values indicated in Table 6.3, which suggest that pitch salience was sufficient for discrimination when CT were masked, but not when they were audible, for f_c=7 kHz. Such an advantageous effect of the background noise could be explained by mechanisms of spectral completion, which may have allowed the listeners to infer the presence of lower stimulus components than those effectively present (McDermott and Oxenham, 2008). Therefore, when all stimulus components lie above 5 kHz, the use of masking noise to eliminate cochlear distortion products may paradoxically reintroduce subjective components at the exact same frequencies at a central perceptual level, with important consequences for the salience of the perceived pitch. The use of background noise in the present work might thus account for the finding of a salient pitch when all stimulus components lay above 5 kHz, *i.e.*, higher than the “existence region” of residue pitch, as measured by Ritsma (1962) in quiet. In fact, Oxenham *et al.* (2011) recently showed that this “existence region” could be extended to higher spectral regions, using high-frequency complex tones presented in background noise.

Finally, the fact that d' values were always above 2 in all masked-CT conditions (Table 6.3) implies that listeners can be expected to perform largely above chance in H/I-discrimination experiments when center-component ranks below 11.5 are used, independently of whether combination tones are properly masked or not. In other words, performance in the H/I-discrimination experiments of Moore *et al.* (2009b) and Moore and Sęk (2009a) should not crucially depend on adequate masking of combination tones, as long as the physical components of the stimulus remain above masking threshold, and as long as a center harmonic rank of 11 is used. However, for higher values of N, adequate CT masking may more critically impede performance, all the more so as higher CT levels are expected as N increases (*e.g.*, Goldstein, 1967; Hall, 1972).

A set of two-sample Kolmogorov-Smirnov tests performed on the individual data sets for each condition confirmed that the presence of audible CT gave rise to significantly different distributions of matches ($p<0.05$) in 27 out of 36 comparisons. Most of the cases where no significant difference was found corresponded to f_c=7 kHz, where more erratic matches were obtained due to a lower pitch salience.

In summary, the audibility of combination tones was found to have an influence on the perceived pitch, and on pitch preference among several ambiguous pitches. However, it generally did not affect pitch salience, which was estimated sufficient for accurate H/I-discrimination when all CT were adequately masked. Whether the observed effects can be accounted for by different pitch mechanisms will be discussed in Chapter 7.

As a final remark unrelated to effects of phase or CT audibility, one may note the more balanced mixing proportions between distributions below and above f_{env} in the present study (Fig. 6.4, masked-CT conditions), compared to those obtained by Santurette and Dau (2011a). In their study, the listeners showed an overall preference for the higher pitch. Such a difference is consistent with the fact that they used a value of N slightly below 11.5, whereas N was exactly equal to 11.5 in the present study, which corresponds to the most ambiguous situation with an "inharmonicity index" of 0.5 (de Boer, 1956a).

6.7 Experiment 5: Influence of dichotic presentation on pitch matches

This pitch-matching experiment investigated the effect of presenting every other stimulus component to the opposite ear on the low pitch. Such a dichotic presentation mode was previously used by Houtsma and Goldstein (1972), who found no difference between monotic and dichotic presentation for performance in musical interval recognition. This finding indicated that the peripheral interaction of components was not necessary for complex pitch perception and that pitch mechanisms operated centrally, based on inputs of the same nature, whether these resulted from monotic or dichotic stimulation. Using a similar approach, Bernstein and Oxenham (2003) found that, for 12-component complex tones with F0=[100, 200] Hz, dichotic presentation elicited a pitch at F0 when harmonics below the 10th were present, whereas a pitch at 2F0 was heard if the lowest harmonic rank was 15 or higher. This experiment investigated the effect of dichotic presentation on the 5-component inharmonic complex tones of interest here. Four listeners (subjects 6 to 9) participated in this part of the study.

6.7.1 Method

The stimuli and procedure were the same as in Experiment 4, except that components at f_c-2f_{env}, f_c, and f_c+2f_{env} were presented to the left ear, while components at f_c-f_{env} and f_c+f_{env} were presented to the right ear. All components had a starting phase of 0. The same background noise as in the “masked CT” conditions of Experiment 4 was presented diotically, and the matching pulse trains were presented monaurally in each listener’s best ear.

6.7.2 Results and discussion

The distributions of matches for the whole subject group (160 matches per condition) are illustrated in Fig. 6.5, together with the fitted Gaussian mixture models. The corresponding model parameters are detailed in Table 6.4.

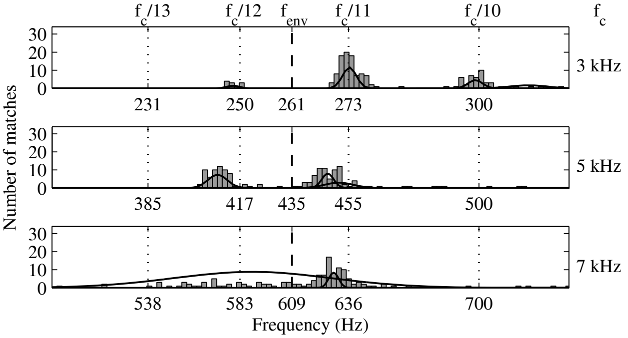


Figure 6.5: Pitch matching of the fundamental frequency of broadband pulse trains (horizontal axis) to five-component high-frequency complex tones with a center component at f_c and an envelope repetition rate $f_{env}=f_c/11.5$, for dichotic presentation and masked combination tones. See the caption of Fig. 6.4 for more details.

Table 6.4: Best fitting Gaussian mixture model parameters to the overall pitch matching results of Experiment 5. For each condition, the mean and standard deviation (in brackets) of each distribution are listed, as well as the proportion of the mixture represented by that distribution. The predicted d' values for a discrimination task are also listed at the bottom of each column. Stars indicate d' values significantly above chance level for a task similar to that of Experiment 1.

f_c	3 kHz		5 kHz		7 kHz	
1	248.6	(1.3)	408.9	(3.4)	589.4	(38.5)
		6.9%		36.5%		44.2%
2	272.9	(2.1)	447.3	(4.3)	629.1	(6.0)
		58.0%		39.8%		42.1%
3	299.2	(2.2)	451.2	(25.5)	786.8	(70.8)
		23.0%		14.7%		13.7%
4	310.3	(15.2)	564.1	(34.2)		
		8.5%		8.9%		
5	368.0	(6.6)				
		3.1%				
6	418.5	(0.0)				
		0.6%				
d'	5.6*		3.9*		2.5*	

As can be observed by comparing the distributions of matches in Fig. 6.5 to those in Fig. 6.4(a) and 6.4(b), there was no clear effect of dichotic presentation on the low pitch. The statistical analysis (Table 6.4) further revealed similar distribution means for the dichotic and monaural conditions, as well as the absence of matches around $2f_{env}$, which would have been expected if the pitch had relied on independent ISI information from the left and right peripheral channels. Pitch salience was similar for the monaural and dichotic conditions, in which it was overall slightly more accurate, as reflected by the standard deviations of the fitted distributions and the estimated d' values (Table 6.4). The three listeners who had also participated in Experiment 4 (subjects 6, 7, and 9) all showed a slight increase of the proportion of matches above f_{env} in the dichotic condition, compared to the monaural conditions. Subject 9 reported he sometimes heard two different pitches in the left and right ear. However, this was not reflected in his matches, which always corresponded to a combined percept from both ears.

The absence of a clear difference in the low pitch for monaural *vs.* dichotic presentation is consistent with the use of spectral cues. However, a temporal autocorrelation mechanism may also be able to account for the present results (Bernstein and Oxenham, 2003). These aspects will be discussed in Chapter 7.

6.8 Experiment 6: Ability of the listeners to hear out individual components

The lack of phase effects in Experiment 4 suggested that the components of the complex tones may not have been completely unresolved. Moreover, no evidence for the use of timing information was found in Experiment 5. In this last experiment, component resolvability was thus evaluated directly by testing whether the listeners were able to hear out the three lowest spectral components, using a method similar to that described by Bernstein and Oxenham (2003). The procedure was slightly modified compared to that used in the similar experiment of Santurette and Dau (2011a). In particular, it was ensured here that each individual listener was provided sufficient training with similar stimuli to those used in the measurement runs. Six subjects participated in this experiment.

6.8.1 Method

The task of the subjects was to identify which of a comparison or a target tone was higher in frequency. A 2I-2AFC procedure was used. In each trial, two 1-s intervals separated by a 375-ms silent gap were presented. The first interval contained three bursts of a 300-ms sinusoidal comparison tone with frequency f_{comp}, each including 20-ms onset and offset cosine ramps, separated by 50-ms silent gaps. The second interval contained a 1-s complex tone, in which the target component with frequency f_{targ} was first filtered out. This target component was then gated on and off in the same way as in the first interval. The complex tones had identical component amplitudes to those used in Experiments 2, 4, and 5, but all components were generated with random starting phase. The comparison and target tones were both presented at the same level as that of the corresponding component in the original complex tone. No background noise was present in this experiment. In each trial, f_{comp} was either lower or higher than f_{targ}, with equal probability, and the absolute frequency difference between f_{comp} and f_{targ} was chosen from a uniform distribution of values between $0.035f_{targ}$ and $0.05f_{targ}$. In order to prevent the subjects from using memorization cues, the center frequency of the complex f_c was roved between $0.935f_c$ and $1.065f_c$, and all conditions were presented in a random order within one run. Each run contained 30 trials for each of 9 conditions (3 target components for each of the 3 f_c values), and the last 25 trials were included in the results. In each run, the N parameter was fixed, and the first five trials for each condition were not included in the final results. Each subject first performed one run for N=5.5 and one for N=8.5, in both of which feedback was provided. Two runs for N=11.5 were then performed, in which feedback was not provided. All listeners performed training runs with N=5.5 until their performance reached 90% correct in at least one condition.

6.8.2 Results and discussion

The average results and standard deviations over all subjects are plotted in Fig. 6.6 as a function of f_c and the rank $n=f_{targ}/f_{env}$ of the target component. For a given condition, a star indicates that the mean score was significantly above chance level (68% correct required for significance for N=[5.5, 8.5], 60% for N=11.5). Given the large across-subject variability, the conditions in which individual scores were significantly above chance level are also marked with subject numbers.

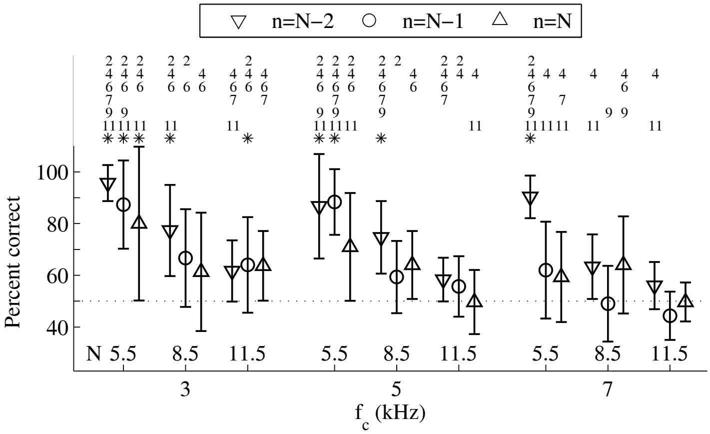


Figure 6.6: Ability of six subjects (means and standard deviations) to hear out the three lowest components of 5-component complex tones, as a function of the rank N and frequency f_c of the center component. $n=f_{targ}/f_{env}$ indicates the rank of the target component. Subject numbers indicate individual percent correct scores significantly above chance level (stars for mean scores).

Overall, performance worsened with increasing f_c and increasing n. A within-subjects two-way ANOVA confirmed significant effects of both f_c [$F(2,135)=11.14$, $p<0.0001$] and n [$F(8,135)=9.74$, $p=0.0001$], and there was no interaction between the two factors [$F(16,135)=0.78$, $p=0.71$]. For f_c=[3, 5] kHz, performance remained significantly above chance in a majority of subjects up to n=6.5. For $n\geq7.5$, the average scores approached chance level. However, a few listeners still scored significantly above chance for some conditions where $n\geq7.5$, especially for low f_c values. This contrasts with the results of Santurette and Dau (2011a), whose listeners' performance did overall not raise significantly above chance level for any f_c value in a similar experiment. However, given the across-subject standard deviations reported in both studies, the group data do actually not differ as a whole. The improvement in performance with decreasing f_c observed here was nevertheless not reflected in the results of Santurette and Dau (2011a). The fact that no background noise was used and that the listeners were provided stimulus-specific and individualized training in the present study may explain these slightly higher scores for low f_c values, despite the use of roving of f_c and of random component phases. Another explanation for the drop in performance with increasing f_c is the fact that the range of relative differences between f_{comp} and f_{targ} was kept the same for all f_c values in the present experiment. As frequency difference limens for pure tones are known to increase with absolute frequency (Wier, 1977), when expressed as the Weber fraction, the present task may have been more difficult toward high f_c values.

As the group data does not allow a definitive conclusion for N=11.5 and only a subgroup of listeners was able to hear out the target component for each condition, it is of interest to compare the individual results to the pitch-macthing data with identical stimuli. Subjects 2, 6, 7, and 9 had also participated in Experiment 4. The pitch salience, as reflected by the standard deviations of the different clusters of pitch matches, was strongest in subjects 6 and 7, and weakest in subject 9. This is broadly consistent with the performance of these listeners in hearing out individual components, with subject 9 never scoring significantly above chance level for any component. The fact that subjects 6 and 7 could only hear out the lowest component for f_c=5 kHz, whereas both the lowest and center components were heard out for f_c=3 kHz, might also explain why these two listeners showed slight downward pitch shifts from $f_c/12$ for f_c=5 kHz, but not for f_c=3 kHz, even when combination tones were masked. Such observations suggest that both pitch and pitch salience might be determined by which components of the stimuli are resolved, and how well.

However, both subjects 2 and 6 could clearly hear a salient pitch for f_c=7 kHz, despite an inability to hear out any of the stimulus components. This confirms the findings of Santurette and Dau (2011a) that the ability to hear out individual partials is not necessary for a salient low pitch to be evoked. This is in line with the fact that the low pitch sensation arises “automatically”, without an active effort of the listeners, whereas the subjects must focus their attention on the target component to perform the task of Experiment 6, which is cognitively more demanding. Therefore, a definition of resolvability based on the ability of listeners to hear out individual partials does not satisfactorily account for the present pitch matches, should they rely on the presence of resolved components.

6.9 Summary and conclusions

In Experiment 1, discrimination of harmonic and inharmonic, frequency-shifted, 5-component complex tones was found to be significantly above chance for a rank of the center component N=11, but generally not possible for N=15, for F0 between 200 and 600 Hz. A comparison of the results to the data of Moore *et al.* (2009b) suggested that performance in such a task may rely on the rank of the lowest audible component present in the stimulus. The discrimination abilities of the listeners were in line with pitch-matching data, suggesting that H/I-discrimination relies on a pitch cue. Similar trends in the H/I-discrimination and pitch-matching results suggested that combination tones might have been audible in the high-frequency conditions.

The results of Experiment 2 revealed that the level of the most prominent combination tone was as high as that of the lowest stimulus component in some listeners. Such a level was lower for stimulus components in ALT phase than in SIN phase. In Experiment 3, it was shown that the background-noise level used in Experiment 1 was sufficient to mask combination tones for $f_c \leq 5$ kHz, but that combination tones might have been audible in some listeners for f_c=7 kHz.

The pitch matches obtained in Experiment 4 for a center-component rank N=11.5 showed no effect of relative component phases (SIN or ALT) on the perceived pitch, indicating that the stimulus components were not completely unresolved in terms of phase effects. Therefore, no evidence for the use of temporal cues for pitch extraction was found here, even though such a result cannot rule out the use of temporal information. The audibility of combination tones was found to produce a pitch shift in most listeners, and to affect pitch preference between several ambiguous pitches. However, it did overall not affect pitch salience, suggesting a limited role of combination-tone audibility for H/I discrimination tasks in which the center-component rank is 11 or below. For such ranks, performance in such tasks is indeed expected to be significantly above chance even when combination tones are properly masked. If pitch salience was not affected by CT-audibility, it was however found to be reduced in the absence of background noise for f_c=7 kHz. This suggests that the listeners might use compensation strategies for perceiving high-frequency complex pitch by inferring the presence of a plausible signal in the noise, when all stimulus components lie above 5 kHz.

In Experiment 5, it was found that presenting every other stimulus component in the opposite ear did not affect the low pitch, compared to monaural presentation of all components as in Experiment 4. Here again, an effect of dichotic stimulation would have indicated the use of temporal cues, but the absence of such an effect does not allow any clear conclusion on the nature of the underlying mechanisms.

Experiment 6 investigated the ability of the listeners to hear out the individual stimulus components. The trends in the individual results as a function of target-component rank and absolute frequency were consistent with the pitch matches obtained in Experiment 4. This could suggest a link between the accuracy of the representation of individual partials at the cochlear output and the perceived pitch. However, the extent to which this accuracy is sufficient for pitch extraction was not correctly measured by the present task. In other words, the partials may be have been resolved enough for pitch perception, but unresolved in terms of the ability of listeners to hear them out.

This last point raises the question of an adequate definition of resolvability. In recent H/I-discrimination (Moore *et al.*, 2009b; Moore and Sęk, 2009a) and pitch-matching (Santurette and Dau, 2011a) studies, the assumption of temporal pitch cues was made on the basis of unresolved partials. However, the components were considered unresolved as long as the listeners could not hear them out from the complex. The present results suggest that this may not apply to complex pitch perception.

It should be emphasized that the present findings do not rule out the use of temporal fine-structure cues for the low pitch of high-frequency complex tones with “intermediate” component ranks such as the ones used here. However, they are inasmuch consistent with the use of spectral cues, as no evidence for temporal cues was found in the above experiments. In an attempt to determine the extent to which the use of spectral and temporal information can account for the present results, pitch predictions will be obtained using either or both types of information in the following chapter.

Acknowledgments

Parts of this work were carried out at the Auditory Perception and Cognition Lab, Department of Psychology, University of Minnesota, Minneapolis, USA. I would like to thank Christophe Micheyl for scientific discussions and for his help concerning the prediction of discrimination performance from the pitch matching data, Andrew J. Oxenham for fruitful discussions, and Brian C.J. Moore for supplying the program for excitation-pattern calculations.

7

Implications for pitch mechanisms

In this chapter, it was attempted to determine which of three suggested neural pitch representations of the stimuli was most likely to account for the pitch matches obtained in Chapter 6. This was achieved by obtaining physiologically realistic spatiotemporal auditory-nerve activity patterns from a peripheral model, and using either the spectral information, the temporal information, or both types of information contained in such patterns, for pitch estimation. The model simulations and the different paradigms used for pitch extraction were kept simple, as the aim here was more to compare the likelihood of the use of specific types of information in the light of the present data, rather than to find a model which would account in details for all psychophysical observations.

7.1 Model simulations

7.1.1 Peripheral auditory model

In order to investigate whether the pitch-matching results reported in Chapter 6 could be accounted for by different pitch theories, internal spatiotemporal representations of the stimuli at the level of the auditory nerve were obtained *via* the peripheral auditory model of Zhang *et al.* (2001). Such a model was chosen because it provided physiologically realistic outputs in terms of both firing rate and timing as a function of the characteristic frequency (CF) of high-spontaneous-rate fibers. This made it possible to compare pitch predictions based on mechanisms using either rate-place information, interspike-interval (ISI) information, or operating directly on the two-dimensional spatiotemporal activity pattern. Such patterns were obtained by feeding 200-ms samples of the complex tones used in Chapter 6 to the model, with a sampling rate of 48 kHz and an input stimulus level of 50 dB SPL. The most prominent combination tone (CT) was either omitted or added using the levels and phases obtained by subject 6 in Experiment 2 (see Table 6.1). The starting phase of the CT was in the latter case assumed to be in opposite phase to the measured cancellation phase. The model parameters for the healthy human ear were used, with a spontaneous firing rate of 50 spikes/s. The spatiotemporal activity patterns were derived from the responses at the synapse output, obtained in 400 frequency channels with CFs ranging from 1 to 10 kHz.

7.1.2 Internal pitch representations

Three internal profiles were derived from the output of the peripheral model:

- A rate-place profile, based on the amount of activity along the tonotopic axis, corresponding to a spectral representation;
- A mean-absolute-spatial-derivative (MASD) profile, based on the relative timing of the activity in neighboring frequency channels, corresponding to a spectrotemporal representation;
- An ISI profile, based on the autocorrelation functions of the activity in individual frequency channels, corresponding to a temporal representation.

How these profiles were obtained and used for pitch predictions is described in the following[1].

---

[1] For comparison purposes, simulations were also obtained using a simple gammatone filterbank (Patterson *et al.*, 1995) as the only peripheral processing stage. The output profiles were found to be very similar to those reported here. This is hardly surprising, because the Zhang *et al.* (2001) model is also based on gammatone filters. The only notable difference obtained with their model is due to the inclusion of the rapid saturation of the spike rate with increasing level, leading to a smaller dynamic range in the rate-place profile.

Rate-place profile

The spike rate as a function of CF was used to obtain pitch predictions based on a spectral mechanism. Subharmonics of CFs corresponding to channels in which the spike-rate showed a maximum were used to build a Schroeder histogram (Schroeder, 1968). A binwidth of 3 Hz was used, as it corresponded roughly to the mean pitch accuracy of the listeners who participated in the pitch-matching experiments (*cf.* Table 5.1). Histograms obtained for different bin center-frequencies, in steps of 0.1 Hz, were averaged. All peaks in the rate-place profile and all corresponding subharmonics were given the same weight in the histogram calculation.

MASD profile

The MASD profile relies on a spectrotemporal operation that emphasizes the phase-transition cues created by the cochlear traveling wave (Cedolin and Delgutte, 2010), by assuming a lateral-inhibition mechanism (Shamma, 1985). It is obtained by calculating the derivative of the activity pattern along the CF dimension, then integrating its absolute value over time. The result is a one-dimensional profile, as a function of CF. The MASD profiles were obtained here using discrete derivation and integration operations, and pitch predictions were derived in the same way as for the rate-place profile.

ISI profile

In order to derive the ISI profile, autocorrelation was performed independently for each CF, and all temporal outputs were then summed across channels to obtain a summary autocorrelation function (SACF) (Meddis and Hewitt, 1991). The inverse of the maximal lags near $1/f_{env}$ in the SACF, where f_{env} is the envelope repetition rate of the stimulus, were used as pitch predictions. The autocorrelation functions were obtained from responses at the synapse output, and do thus not stricly reflect ISI histograms. However, performing the analysis on the model output after spike generation, *i.e.*, using discharge times generated by a Poisson process, did not give rise to different SACFs on average. For the simple pitch predictions intended here, the spike-generation stage of the model was thus omitted.

7.2 Pitch predictions

It should be emphasized here that the pitch-extraction algorithms described above were kept simple on purpose, and are not pretended to be the most plausible physiologically or to account for most aspects of pitch perception described in the literature in the best way. For instance, the equal-weighted histogram is an arbitrary choice among many other options when interpreting a tonotopic activity profile, and different weights may better reflect known psychophysical trends. In the same way, weights may be attributed to different lags in the SACF. Specific orders of ISIs, as well as specific frequency channels, may also convey more information than others, and thus also deserve different weights in an internal time-interval representation. Besides, the MASD profile chosen here is obtained by comparing timing information in neighboring channels only, and across-channel comparisons between more distant channels may also occur. These issues are important to consider and have been raised in the literature, leading to new or improved pitch-perception models. However, they are out of the scope of the present thesis, and the simple choices made here were thought sufficient to evaluate the potential use of different types of information and address the following questions.

For the specific stimuli used in Chapter 6, are simple pitch-extraction schemes based on each of the three hypothetical internal profiles described above able to:

- accurately predict the different ambiguous pitches?
- predict the absence of phase effects on the perceived pitch?
- predict a pitch shift in the presence of audible CT?
- account for changes in pitch preference between the different conditions?
- predict similar pitches for monotic and dichotic presentation?

7.2.1 Influence of relative component phases and combination-tone audibility

The internal profiles and the corresponding pitch predictions were obtained for the 5-component complex tones used in Chapter 6 (see section 6.6), with a frequency f_c=5 kHz and a rank N=11.5 for the center component. The background noise was not included in the simulations[2]. The obtained rate-place (solid lines) and MASD (dashed lines) profiles are plotted in Fig. 7.1. These are given for the SIN (left panels) and ALT (right panels) phase configurations. In addition, simulations in which no CT was added (upper panels) and in which the estimated most prominent CT was added to the stimulus (lower panels) are shown. The obtained SACF profiles are plotted in Fig. 7.2. All three profiles are normalized with respect to their maximum value.

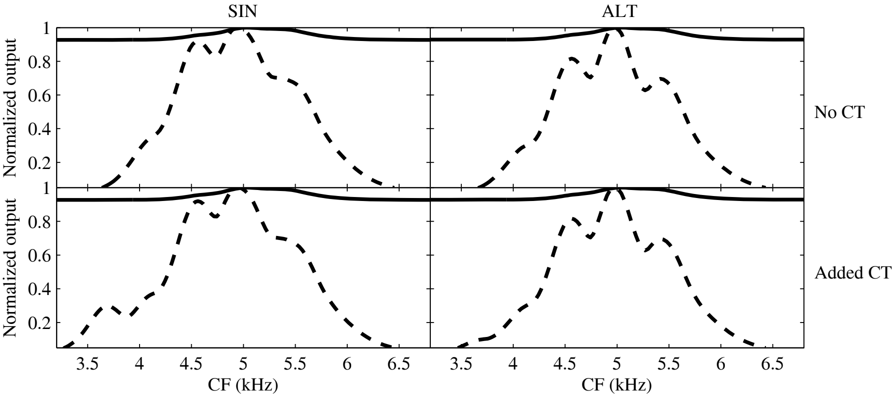


Figure 7.1: Rate-place (solid lines) and MASD (dashed lines) profiles for the 5-component complex tones used in Experiment 4 (see section 6.6), with f_c=5 kHz and N=11.5. Four profiles are plotted: for components in SIN phase (left panels) and ALT phase (right panels); for simulations in which the most combination tone is omitted (upper panels) and added (lower panels).

In Fig. 7.1, it can be seen that the rate-place profiles (solid lines) show little

---

[2] Simple simulations indicated that the average model output when the background noise was included did not lead to different peak locations in any of the three profiles. The two main effects of the noise were an increase in activity at low CFs in the MASD profile, and a loss of resolution in the SACF due to the random temporal fluctuations, especially if only few averages were used.

dependence on CF, due to the fact that the firing rate partly reaches saturation at the moderate sound level used here (50 dB SPL). Thus, according to a rate-place representation, the individual components of the stimulus are not well resolved. This is consistent with an excitation-pattern representation (*e.g.*, Moore and Glasberg, 1997) which does not show well-defined ripples for unresolved components. The obtained rate-place profiles are very similar for SIN and ALT phase configurations, which is consistent with an absence of phase effects. However, the rate-place profiles are not largely affected by CT audibility, and are thus expected not to predict the pitch shifts caused by audible combination tones.

In contrast to the rate-place profiles, the MASD profiles (dashed lines) show much more defined contours, with maxima occuring at CFs corresponding roughly to the frequencies of the stimulus components with the largest amplitude. The simulated lateral-inhibition process thus greatly emphasizes the internal spectral representation of the stimuli. In terms of the MASD profile, some components may indeed be considered resolved. When comparing the MASD profiles for stimuli in SIN *vs.* ALT phase, it can be observed that the location of the peaks in the profile is roughly similar, consistent with an absence of phase effects on the perceived pitch. However, the relative amplitude of the activity peaks differs, which may account for a change in pitch preference between several ambiguous pitches. In the pitch-matching experiment, the listeners tended to prefer a higher pitch in the ALT configuration than in the SIN configuration (section 6.6.2). The obtained MASD profiles might be able to reflect this by the presence of better-resolved peaks towards higher CFs for ALT-phase stimuli in comparison to SIN-phase stimuli. Finally, the effect of adding the most prominent CT is clearly visible in the MASD profile, in which it creates a peak around the CT frequency. Therefore, unlike the rate-place profile, the MASD profile is expected to allow the prediction of a pitch shift when the CT is added. Note that in Fig. 7.1, the difference in activity in the MASD profile around the CT frequency between the SIN and ALT phase configurations reflects a lower CT level for components in ALT phase.

The obtained SACFs (Fig. 7.2) show several maxima near the lag corresponding to f_{env} (2.3 ms), which may account for the perceived ambiguous pitches. However,

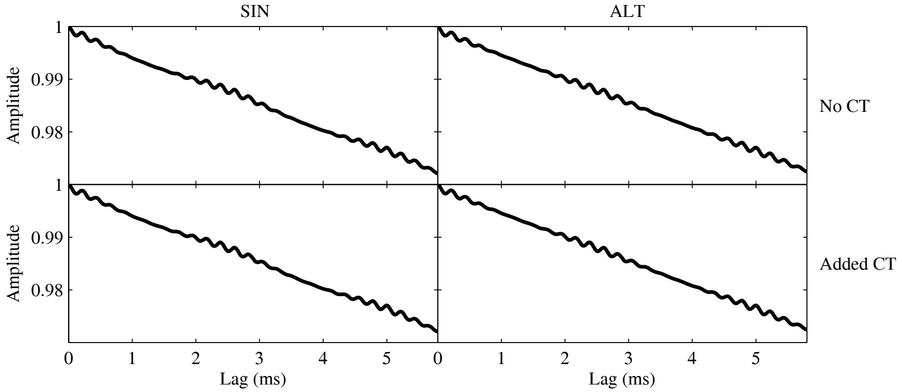


Figure 7.2: SACFs for the 5-component complex tones used in Experiment 4 (see section 6.6), with f_c=5 kHz and N=11.5. Four profiles are plotted: for components in SIN phase (left panels) and ALT phase (right panels); for simulations in which the most combination tone is omitted (upper panels) and added (lower panels).

there is no observable difference between the SACFs in the different conditions. This suggests that such an ISI profile may be able to account for a lack of phase effects, but not for an effect of CT audibility.

The pitch predictions obtained from the rate-place, MASD, and ISI profiles shown in Fig. 7.1 and 7.2 are summarized in Fig. 7.3. For each condition, Schroeder-histogram predictions are shown as bold solid lines for the rate-place profile and as bold dashed lines for the MASD profile. Circles indicate the pitch predictions from maxima in the SACF, corresponding to the ISI profile.

In all cases, an ambiguous pitch was predicted, which never corresponded to the envelope repetition rate, and was close to subharmonics of f_c in the absence of CT. This is in line with the behavioral data reported in Fig. 6.4. The rate-place and ISI profiles generally led to similar predictions. In comparison, using the MASD profile led to slightly lower predicted pitches, especially for pitches below f_{env} (compare histogram peaks for the solid *vs.* dashed lines in Fig. 7.3). As human listeners generally showed pitch matches slightly below subharmonics of f_c in Chapter 6 (see Fig. 6.4), predictions obtained from the MASD profile are thus closer to the psychophysical data[3].

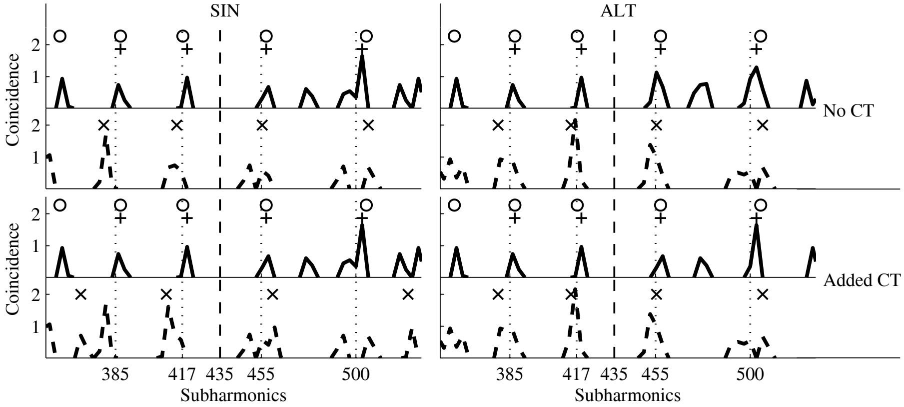


Figure 7.3: Pitch predictions for 5-component complex tones with f_c=5 kHz and N=11.5, for stimuli in the SIN (left panels) and ALT (right panels) phase configurations, and with absent (upper panels) or present (lower panels) CT. For each condition, Schroeder-histogram predictions are given based on the rate-place profile (bold solid lines) and on the MASD profile (bold dashed lines). Predictions based on the lowest effective component are indicated with '+' signs for the rate-place profile and '×' signs for the MASD profile. Circles indicate the inverse of maximal lags in the SACF. The fine dashed lines indicate f_{env}, and the fine dotted lines subharmonics of f_c.

As a comparison to the equal-weighted Schroeder histogram, pitch predictions were also obtained for the assumption of a pitch determined exclusively by the lowest effective component. In this case, only the lowest CF giving rise to a peak in the rate-place or MASD profile was used, and the predicted pitch was assumed to correspond to a subharmonic of this specific CF (Walliser, 1969). The comparison of the equal-weight-histogram (EWH) and the lowest-effective-component (LEC) predictions to the experimental data aimed to determine whether the lowest components dominated the pitch percept. In Fig. 7.3, the LEC predictions are indicated with '+' symbols for the rate-place profile and with '×' symbols for the MASD profile. Here again,

---

[3] Some caution is needed when interpreting small differences between the different pitch predictions. Specifically, it should be kept in mind that the uncertainty around the position of histogram peaks depends on the binwidth used (3 Hz in the present case) and on the number of frequency channels in the model (with 400 channels, neighboring CFs around 5 kHz are separated by about 17 Hz).

comparing the position of '+' and '×' symbols indicates the prediction of lower pitches from the MASD profile than from the rate-place profile, for pitches below f_{env}. This makes the predictions from the MASD profile closer to the trends observed in the behavioral data. Overall, little difference was observed between EWH and LEC predictions. However, in the condition for SIN phase and added CT (lower left panel), the LEC prediction from the MASD profile better reflected the lowest perceived pitch than the EWH predictions. This suggests that pitch extraction mechanisms may mostly rely on the information conveyed by the lowest effective components in the stimulus.

All three profiles were able to predict the absence of phase effects on the low pitch. However, only the MASD profile was able to predict a pitch shift with the addition of the most prominent CT. This is mainly visible for the SIN configuration, in which the CT level is the highest. Finally, the SACF alone cannot account for changes in pitch preference between the several ambiguous pitches in the different conditions. Slight differences in the histogram peak-heights in both the rate-place and MASD profile could in principle account for such differences in pitch preference. However, with the simple pitch-estimation scheme used here, histogram peak-height did overall not accurately predict which of the several ambiguous pitches was preferred.

7.2.2 Influence of dichotic presentation

In Chapter 6, presenting alternate components of the complex tones in the opposite ear was found to have little effect on the perceived pitch and on its salience, in comparison to the monaural presentation of all components (see section 6.7). Here, spatiotemporal activity patterns were obtained independently for the input stimuli in the left and right ear.

The corresponding rate-place and MASD profiles are plotted in the upper panels of Fig. 7.4. Such profiles show maxima around CFs corresponding to the frequencies of the stimulus components. The fluctuations in activity as a function of CF are more pronounced than when the five components are presented to the same ear (*cf.* Fig. 7.1), reflecting the higher resolvability of neighboring partials due to the increased component-spacing. Independently, the left and right profiles cannot account for the perceived pitch. However, the similarity in pitch perception for the dichotic and monotic conditions can be accounted for by a combined profile (Fig. 7.4, lower panel), simply derived from the sum of the left and right spatiotemporal activity patterns[4]. Such a profile indeed resembles that obtained for monaural presentation of all components (see Fig. 7.1, upper-left panel).

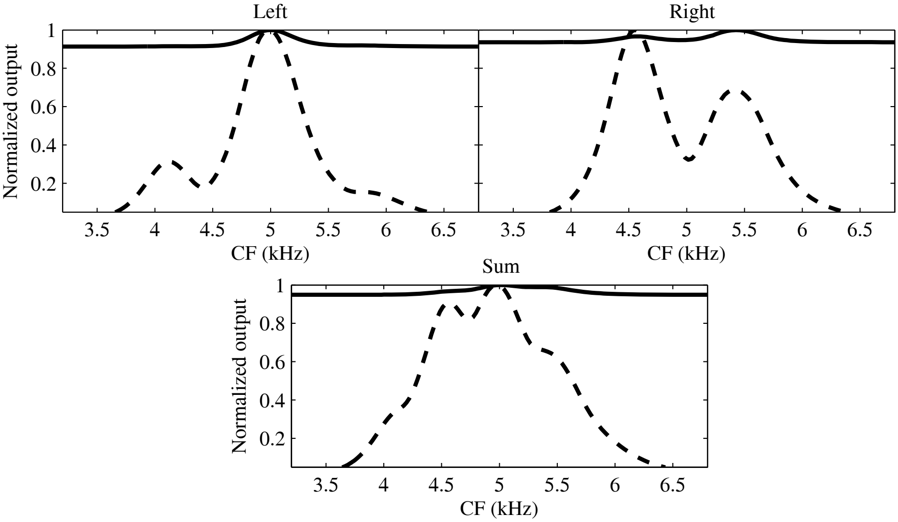


Figure 7.4: Rate-place (solid lines) and MASD (dashed lines) profiles for the dichotic complex tones used in Experiment 5 (see section 6.7), with f_c=5 kHz and N=11.5. The profiles obtained from the independent activity patterns in the left and right ear (upper panels) and from the sum of the left and right activity patterns (lower panel) are plotted.

The SACFs obtained from the left and right spatiotemporal representations are plotted in the upper panels of Fig. 7.5. Here again, these independent ISI profiles cannot account for the perceived pitch. However, the SACF derived from the sum of the left and right spatiotemporal activity patterns[4] (Fig. 7.5, lower panel) resembles that obtained for monaural presentation of all stimulus components (see Fig. 7.2).

---

[4] Note that summing or averaging the left and right output profiles instead of summing the left and right spatiotemporal activity would give rise to similar combined patterns.

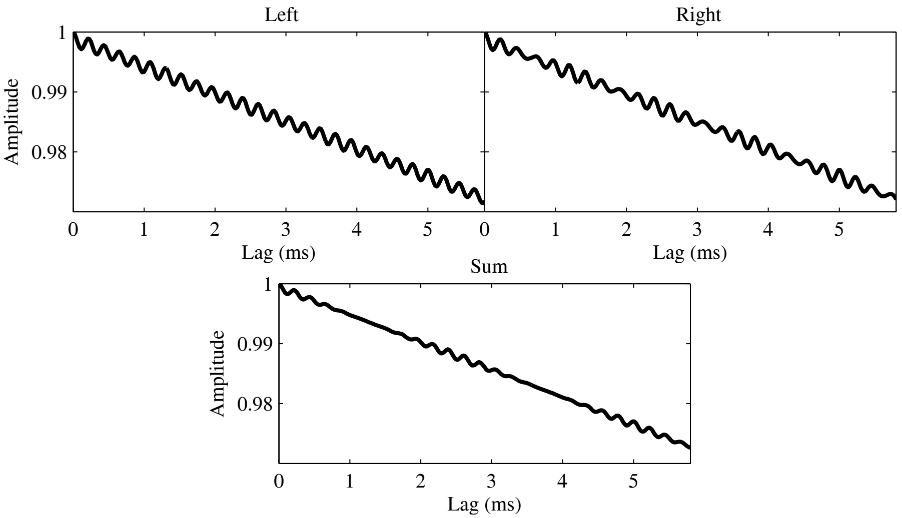


Figure 7.5: SACFs for the dichotic complex tones used in Experiment 5 (see section 6.7), with f_c=5 kHz and N=11.5. The SACFs obtained from the independent activity patterns in the left and right ear (upper panels) and from the sum of the left and right activity patterns (lower panel) are plotted.

The simple simulations shown here indicate that the use of spectral and temporal cues can, in principle, both account for the pitch of the dichotic complexes. This can be seen in Fig. 7.6, where the pitch predictions obtained from the summed left and right profiles are shown. Such predictions are very similar to those for monaural presentation (see Fig. 7.3, upper-left panel). The differences observed between predictions from the three internal profiles also show the same trends to those observed in the monaural conditions, for which predictions from the MASD profile best reflected the psychophysical observations.

7.3 Summary and discussion

Three hypothetical internal pitch representations of the complex tones used in Chapter 6 were derived using a peripheral auditory model: a rate-place profile, an MASD

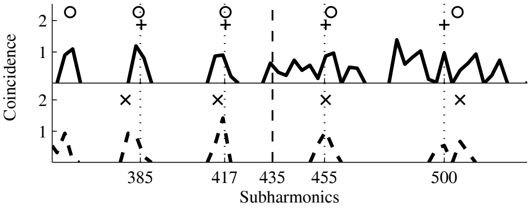


Figure 7.6: Pitch predictions for 5-component complex tones with f_c=5 kHz and N=11.5, presented dichotically. See the caption of Fig. 7.3 for legend details.

profile, and an ISI profile. Simple pitch-extraction algorithms were then used to obtain pitch predictions from these three profiles. For the rate-place and MASD profiles, histograms were built from the subharmonics of CFs showing peaks in tonotopic activity. For the ISI profile, an autocorrelation operation was performed independently in each channel and the outputs were summed across CFs.

It appeared that all three profiles could predict the pitch ambiguity of the complex tones, with pitches near subharmonics of f_c. The absence of a pitch dependence on the relative phase of the stimulus components could also be predicted from all three profiles. However, the pitch predictions from the MASD profile were found to be slightly closer to the means of the distributions of pitch-matches obtained in Chapter 6, compared to predictions from the rate-place and ISI profiles. Moreover, the MASD profile was the only one predicting an effect of audibility of the most prominent CT on the perceived pitch. Therefore, these simple simulations suggest that spectrotemporal mechanisms combining the temporal information across nerve fibers with different CFs might better account for complex pitch perception than purely temporal mechanisms based on periodicity information in individual channels, or purely spectral mechanisms based on tonotopic maxima in firing rate.

As suggested by Cedolin and Delgutte (2010), several weaknesses of the rate-place and ISI representations in terms of reflecting known psychophysical trends could also be overcome by the use of a spectrotemporal mechanism. For instance, the rate-place profile does not degrade toward high frequencies in the same way as pitch salience does, and flattens out rapidly with increasing level, while pitch salience varies to a much lesser extent. Moreover, the ISI profile lacks sensitivity to the effects of component resolvability and component rank on pitch salience. Such limitations are less pronounced for the MASD profile, which is dependent on both spectral and temporal resolution. In Chapter 5, it was also speculated that, in theory, the timing intervals between TFS peaks near adjacent envelope maxima might be retrieved by combining envelope information across frequency channels in the absence of phase locking to the TFS (section 5.4.4). Consistent with this idea, an MASD profile based on the extracted temporal envelopes of the synapse outputs in each channel would show ripples due to the time delays between envelope maxima in neighboring channels. This would lead to the prediction of an ambiguous pitch, whereas an ISI profile based on the temporal information in individual channels would wrongly predict a single pitch close to the envelope repetition rate.

Finally, the above simulations suggested that the absence of an effect of dichotic presentation on the preceived pitch did not rule out the use of a specific type of information. Combining the left and right spatiotemporal activity patterns evoked by the dichotic stimuli indeed led to similar pitch predictions to those obtained in the monaural case. This was demonstrated *via* a basic summation operation on the left and right activity patterns, which is unlikely to reflect the processing of binaural information in the auditory system. However, it remains possible that the peripheral signals are combined in such a way that both spike-rate and spike-timing information at the input of central pitch-extraction mechanisms are indistinguishable for dichotic *vs.* monotic stimulation. One argument favoring the use of place information was raised by Bernstein and Oxenham (2003) and Oxenham *et al.* (2011), who argued that the precision of timing information may be degraded for the dichotic stimuli compared to their monotic counterparts. This is because the temporal envelope at the output of cochlear filters shows smaller dynamic fluctuations in the dichotic conditions, due to less interaction between neighboring partials. According to this view, if pitch relies on temporal cues, dichotic presentation should lead to a lower pitch salience than monotic presentation. This does not correspond to the behavioral observations, which showed a similar pitch salience for the two presentation modes. A temporal code may thus seem less likely. However, such an argument relies on the assumption that the shape of the temporal envelope in the periphery affects phase-locked representations to a sufficient extent to prevent the use of cues based on TFS peaks close to adjacent envelope maxima. Moreover, the extent to which humans are able to use residual phase-locked information at high frequencies remains unknown. It has for instance been argued that the amount of temporal information in the auditory nerve may be sufficient for frequency discrimination up to at least 10 kHz, and that the use of ISI information at such high frequencies is more consistent with behavioral trends than the use of rate-place information (Heinz *et al.*, 2001).

At this stage, it is thus not possible to exclude a role of either place or timing information. One can only underline the limitations of a code based on either purely spectral or purely temporal cues, and the fact that combining both types of information might better account for the low pitch of high-frequency complex tones with intermediate component ranks, such as those considered in the present work.

8

General discussion

In this thesis, several aspects of pitch perception in the human auditory system were investigated using psychophysical methods. The present approach involved the use of artificial stimuli with specific spectrotemporal properties, such as binaural-pitch and transposed stimuli, in an attempt to determine the nature of the underlying mechanisms. The central question of the relative importance of spectral and temporal information for pitch perception was addressed. Another aim was to relate binaural pitch perception to specific auditory functions. This was achieved by studying the effects of hearing impairment on binaural pitch perception outcomes, and comparing these to the performance of individual listeners in basic auditory tasks. Thereby, it was also intended here to help better characterize individual hearing loss and better understand the relationship between the specific deficits observed in hearing-impaired (HI) listeners.

The results reported in Chapter 2 indicated that hearing impairment affected the ability of some listeners to detect binaural-pitch stimuli. A clear distinction was observed between HI listeners for whom binaural pitch perception remained immediate, and HI listeners who could not perceive binaural pitch at all. The listeners with absent binaural pitch perception were the ones who reported deficits in central auditory processing. In listeners who could perceive binaural pitch, hearing impairment was found to affect pitch discrimination of Huggins'-pitch stimuli, as well as melody recognition with different binaural-pitch types. Melody-recognition abilities were found to depend on the interaural phase configuration used in the binaural-pitch stimulus. The highest performance was obtained for stimuli of the Huggins'-pitch type, while the use of a single coherence edge in the stimulus generally led to the poorest performance. Deficits in frequency discrimination were found to partly account for the ability of the listeners to recognize melodies. Interestingly, neither of these two measures were correlated with the listeners' hearing thresholds at low frequencies. This suggested that supra-threshold deficits may be mainly responsible for impaired binaural pitch perception. This was confirmed by simulations of a loss of audibility in normal-hearing (NH) listeners, which did not affect melody recognition with binaural-pitch stimuli. However, simulating a broadening of the auditory filters caused the melody-recognition scores to drop, suggesting a role of frequency selectivity in accurate binaural pitch perception. Despite this, estimates of auditory filter bandwidths in the individual listeners were not correlated with melody recognition scores, suggesting that other factors than frequency selectivity may be crucial for perceiving binaural pitch. Finally, the temporal resolution of the listeners, as measured by a gap-detection task, was found to be correlated with binaural-pitch-perception outcomes, although reduced audibility may influence performance in such a task.

The findings of Chapter 2 suggested that HI listeners could generally be divided into two distinct groups: listeners with immediate binaural pitch perception, and listeners for whom binaural pitch perception is totally absent. They further suggested that the second group might be restricted to listeners with central auditory-processing deficits, in addition or not to impaired peripheral mechanisms. An important role of frequency selectivity was pointed out, but reduced frequency selectivity alone could not account for impaired perception of binaural-pitch stimuli. Following this, Chapter 3 aimed at investigating in more details whether binaural-pitch stimuli may be useful to identify the site of impairment in sensorineural hearing loss, and whether their perception could be directly related to a specific auditory function. Pitch-contour-identification scores were compared for Huggins'-pitch stimuli and monaurally-detectable pitches embedded in noise. The same group distinction as in Chapter 2 was observed in HI listeners. The fact that listeners unable to hear binaural pitch could clearly perceive the monaurally-detectable pitch indicated that their deficit did not stem from a general difficulty in extracting tonal objects from noise. The results of common audiometric measures confirmed that a retrocochlear deficit was very likely in these listeners. However, such deficits may also have been present in some listeners with immediate binaural pitch perception. Therefore, further investigations in listeners with specific clinical diagnoses are needed to determine whether binaural-pitch stimuli may be used in a site-of-lesion test. Deficits in binaural pitch perception were found not to rely on cognitive abilities, but to coexist with a loss of binaural-processing advantage in noise. Such a loss was related to a more basic deficit in temporal-fine-structure (TFS) processing. However, it could not be determined whether the impaired TFS representations responsible for a loss of binaural pitch percept arose centrally, *i.e.*, at or after binaural-integration stages, or from temporal mechanisms in the periphery.

In Chapter 3, it was also attempted to characterize individual "auditory profiles", and a test battery was designed which included measures of basic auditory functions. In addition to the traditional audiometric tests, measures of loudness perception, cognitive function, binaural masking release, TFS processing, and frequency selectivity, were performed in all listeners. With the exception of the cognitive measures, sensorineural hearing loss was found to affect performance in all auditory-profile tasks, albeit with notable individual differences. The main finding was that of an important role of monaural TFS processing for most other tasks. Furthermore, the observed deficits in TFS processing were found to be independent of frequency selectivity. This is in line with recent evidence for a crucial role of TFS information in forming accurate auditory representations of input sounds such as speech, especially in the presence of background noise (*e.g.*, Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008; Strelcyk and Dau, 2009; Ardoint *et al.*, 2010). An adequate measure of monaural TFS processing is thus important to include in a test battery aimed at characterizing hearing loss. However, a reliable behavioral measure of peripheral supra-threshold deficits in TFS processing at low frequencies is currently lacking. In the present work, the detection of a large interaural phase difference in a single tone as a function of its carrier frequency proved a fast and reliable measure, which nevertheless remains a binaural measure. The apparent correlation between monaural and binaural TFS-processing abilities suggests that it might still be a sensible choice, until a monaural test is found. The other auditory-profile measures used here may be relevant in a laboratory study but too time-consuming in a clinical setting. Therefore, further efforts to design an adequate battery of tests would be beneficial. Ideally, such a battery should allow an evaluation of each subject's auditory profile, in detail, but fast, and without omitting crucial measures, but without redundancy.

The heterogeneity of hearing-loss etiologies and audiometric configurations among HI listeners in the present work impose some great caution when attempting to generalize the implications of the present findings. However, such a diversity in hearing losses may not only be seen as a negative, limiting factor. One of the main challenges when carrying out psychophysical experiments with HI listeners is indeed to make use of the large variability often found among subjects. These large individual differences are sometimes blamed for leading to poor correlations when one wishes to compare different measures. This can sometimes hinder the researcher's chances to lead a conclusive study. In order to prevent a large intersubjective variability and reduce individual differences in the results, it is sometimes tempting to gather subjects with similar audiograms, same origin of hearing loss, or other common features. Forming a more homogeneous group in such ways makes it easier to draw conclusions based on averaged results over a group of HI listeners, and often serves as a justification to do so. However, one must keep in mind that the more one narrows down a selection of subjects, the larger a part of the HI population one automatically excludes from a study that might be relevant to them as well. Moreover, one might question the criteria used to make a group of subjects more homogeneous: Does a measure such as the pure-tone audiogram say enough about a person's hearing loss to be used as a grouping criterion? Taking things further, one could easily wonder whether it is at all sensible to calculate average results over a group of HI listeners. As it seems that a group of identical hearing impairments is never to be found, one might as well be more tolerant of result variability and allow a larger spectrum of subjects to take part in each study. Pointing out individual differences instead of focusing on common traits might after all be much more fruitful, and better represent the reality of what hearing loss is all about, in the broad sense. However, if such an individualized approach is to be used, it is always wise to keep some balance: The group of test subjects still has to be carefully selected and adapted to the research question addressed, and unnecessary bias factors reduced as much as possible. The dangers of drawing general conclusions from a group which is, in our case, never homogeneous, should just be kept in mind.

Chapter 4 of the present thesis was concerned with the ability of listeners with developmental dyslexia to perceive binaural-pitch stimuli. Such a study followed contradictory literature reports concerning an impaired binaural pitch perception in dyslexic listeners. A pitch-contour-identification experiment similar to that conducted in HI listeners in Chapter 3 was carried out in a group of dyslexic listeners and a group of matched controls. All dyslexic listeners were found to clearly perceive the Huggins' pitch stimulus, which led to similar pitch-contour-identification scores to those obtained with a similar-sounding, monaurally-detectable pitch. However, a subgroup of the dyslexic listeners showed some difficulty with pitch-contour identification, independent of the monaural or binaural configuration of the stimulus. Moreover, such a difficulty was correlated with deficits in frequency discrimination. The discrepancies found between the outcomes of earlier studies are thus thought to be related to differences in the nature of the experimental tasks, and not to differences in binaural pitch perception *per se*. Overall, the results favor an influence of auditory memory or auditory attention for the reduced performance of some dyslexic listeners in tasks involving pitch comparisons, rather than a role of impaired low-level auditory mechanisms.

The second part of this thesis dealt with more fundamental aspects of complex pitch perception, with high-frequency stimuli containing closely-spaced components but still evoking a low pitch. Under the assumption that temporal mechanisms were involved in the perception of such a low pitch, due to the unresolvability of the stimulus components, Chapter 5 investigated the nature of such mechanisms. More specifically, the hypotheses of a pitch determined by TFS cues *vs.* temporal-envelope cues were tested. Pitch matches of the fundamental frequency of broadband pulse trains to inharmonic transposed tones with five audible components were obtained. The results demonstrated that the pitch of such transposed stimuli was never determined by the periodicity of their temporal envelope, for carrier frequencies between 3 and 7 kHz, and for ranks of the center component around 11.5 and 14.5. The pitch generally became less salient as the absolute frequency f_c or the rank N of the center component were raised, but was still salient for $f_c \geq 5$ kHz and $N \approx 11.5$. Moreover, when a salient pitch was perceived, the pitch matches clustered around frequencies corresponding to the inverse of the timing between TFS peaks near adjacent envelope maxima in the stimulus waveform. A resolvability experiment also showed that the listeners were unable to hear out the individual components of the stimuli, suggesting that these were unresolved. Therefore, such findings support the hypothesis of a temporally-coded "residue" pitch relying on TFS cues (*e.g.*, de Boer, 1956a; Schouten *et al.*, 1962), even in high spectral regions, for the values of N considered here. They are also consistent with the findings of Moore and Sęk (2009a), who argued that TFS cues may persist up to higher frequencies than previously thought. Such an accurate representation of TFS at high frequencies might be explained by the availability of phase-locked information, or by the recovery of TFS cues from a combination of temporal-envelope cues across auditory-nerve fibers, although it is so far unclear how auditory mechanisms could operate to achieve the latter recovery.

The assumption of a temporally-coded pitch made in Chapter 5 relied on a lack of information about the frequencies of individual stimulus components at the cochlear output, *i.e.*, on the unresolvability of these components by the auditory system. This unresolvability was confirmed by the inability of the listeners to hear out individual partials of the stimuli. However, performance in such a task might not adequately reflect how well the auditory system is able to use differences in activity along the tonotopic axis in terms of pitch perception. Moreover, the amount of background noise required to adequately mask combination tones for the specific stimuli used in Chapter 5 might not have been sufficient. If audible, such combination tones may have been resolved, making the use of spectral information plausible by pitch-extraction mechanisms. Chapter 6 described a set of experiments aiming at clarifying the possibility of a spectral code for 5-component complex tones with N=11.5, similar to those used in Chapter 5. First, the performance of subjects in discrimination of harmonic and frequency-shifted complex tones (H/I-discrimination) suggested that the task relied on a pitch cue, but that combination tones might have been audible for high f_c values in the pitch-matching experiment of Chapter 5. A direct measurement of the most prominent combination tone (CT) confirmed that it had been properly masked for $f_c \leq 5$ kHz, but might have been audible for higher f_c values. A pitch-matching experiment then indicated an effect of CT audibility on the perceived pitch, but not on pitch salience, suggesting that CT-audibility should not crucially affect performance in H/I discrimination tasks with similar stimuli. The configuration of relative starting phases of the stimulus components was found to have no effect on the perceived pitch.

Moreover, presenting alternate components in the opposite ear did not affect the low pitch either. If such effects had been found, they would have indicated the use of temporal pitch cues. However, evidence for such temporal cues was not found, and the present observations are thus consistent with a pitch that is extracted using place information. Despite this, these results do not rule out the use of temporal information, as they could also be accounted for by temporal mechanisms such as autocorrelation. Therefore, it was here neither possible to validate, nor to refute, the hypothesis of a temporally-coded pitch made in Chapter 5.

A definition of resolvability based on the ability of the listeners to hear out individual partials was found not to satisfacorily account for the salience of the low pitch. Specifically, the low pitch may remain salient when individual components are not heard out. Despite this, individual differences between listeners were found to show the same trends in their pitch matches and their ability to hear out partials. This suggests that great care ought to be taken when assuming the unresolvability of partials in terms of pitch perception when they cannot be heard out separately. More generally, this also emphasizes the importance of comparing performance in different tasks in terms of individual results, rather than group results. The participation of only a small number of NH listeners in the studies reported in Chapters 5 and 6, each in only a subset of experiments, imposed some limitations in drawing conclusions from individual results. However, if fast auditory-profile measures can be obtained, as suggested in Chapter 3, they could be relevant in future studies of pitch perception in both NH and HI listeners. For instance, relating the observed individual differences in spectral and temporal resolution to differences in pitch perception outcomes might help reveal the nature of the underlying mechanisms.

Finally, in Chapter 7, it was attempted to determine whether the use of spectral, temporal, or spectrotemporal information could best account for the above psychophysical findings. Specifically, spatiotemporal activity patterns at the level of the auditory nerve were obtained by feeding the same complex tones as those used in Chapter 6 to a model of the peripheral auditory system. Pitch predictions were then derived from three internal profiles based on different types of information: a rate-place profile, an interspike-interval (ISI) profile, and a profile derived from the comparison of timing information in neighboring frequency channels. It was found that all three profiles could, in principle, be used to predict the pitch ambiguity of the inharmonic complex tones as well as the approximate location of the perceived pitches. However, the third profile, which involved across-channel timing comparisons, generally better reflected the psychophysical trends. Spectrotemporal mechanisms may thus be interesting to consider in terms of pitch coding, as they can in theory account for the limitations of the auditory system in terms of both spectral and temporal resolution. Despite this, a clear role of across-channel processing for complex pitch perception has yet to be established, and the question of its necessity remains. A search for the exact nature of hypothetical spectrotemporal mechanisms in the auditory system would also benefit from more physiological evidence. Here, the simulation of a simple lateral-inhibition process was implemented, leading to a contrast-enhancement of the tonotopic activity. Independent of the biological plausibility of such a process, it is worth noting that it could be advantageous to account for the perception of some binaural pitch types (*e.g.*, the binaural coherence edge pitch, *cf.* section 2.5.2), as well as the low pitch of complex tones presented monaurally.

Overall, the first part of this thesis pointed out a specific role of the temporal information conveyed by the TFS for binaural pitch perception. The second part suggested that a role of TFS information was also plausible for pitch perception of high-frequency complex tones. However, arguments were also raised which questioned the necessity of temporal cues for complex pitch perception. A wise conclusion to the present thesis can thus only be a very cautious one. In this respect, the affirmation by which de Boer (1976) ended his inspiring review of many years of research on “residue” pitch is still appropriate today: “We cannot know for sure.”. If the formulation remains valid, the context has undoubtedly changed. Models and experimental techniques have been developed which have provided new insights into human pitch perception. The time theory has also regained some plausibility, although it has certainly not disqualified its historical alternative: place. Until measurement methods are developed which allow an objective estimation of auditory-nerve phase locking in humans, the question of the usability of high-frequency TFS cues for pitch perception might thus remain unresolved.

Bibliography

AFNOR (1977). *Recueil des normes françaises de l'acoustique.* Tome 1: Vocabulaire, NFS30-107. Paris: Association Française de Normalisation.

Akaike H. (1974). A new look at the statistical model identification. *IEEE T. Automat. Contr.*, 19(6):716–723.

Akeroyd M.A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.*, 47(S2):S125–S143.

Akeroyd M.A., Moore B., Moore G.A. (2001). Melody recognition using three types of dichotic-pitch stimulus. *J. Acoust. Soc. Am.*, 110(3, Pt. 1):1498–1504.

Akeroyd M.A., Summerfield A.Q. (2000). The lateralization of simple dichotic pitches. *J. Acoust. Soc. Am.*, 108(1):316–334.

Al-Salim S.C., Kopun J.G., Neely S.T., Jesteadt W., Stiegemann B., Gorga M.P. (2009). Reliability of categorical loudness scaling and its relation to threshold. *Ear Hear.*, 31(4):567–578.

Amitay S., Ahissar M., Nelken I. (2002). Auditory processing deficits in reading disabled adults. *J. Assoc. Res. Otolaryngol.*, 3:302–320.

ANSI (1973). *American national psychoacoustical terminology*, S3.20. New York: American National Standards Institute.

Ardoint M., Sheft S., Fleuriot P., Garnier S., Lorenzi C. (2010). Perception of temporal fine-structure cues in speech with minimal envelope cues for listeners with mild-to-moderate hearing loss. *Int. J. Audiol.*, 49:823–831.

ASA (1960). *Acoustical terminology SI*, 1-1960. New York: American Standards Association.

Attneave F., Olson R.K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *Am. J. Psychol.*, 84:147–166.

Baddeley A., Logie R., Nimmo-Smith I., Brereton N. (1985). Components of fluent reading. *J. Mem. Lang.*, 24:119–131.

Banai K., Ahissar M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiol. Neurootol.*, 9:328–340.

Banai K., Ahissar M. (2006). Auditory processing deficits in dyslexia: Task or stimulus related? *Cereb. Cortex*, 16:1718–1728.

Bernstein J.G., Oxenham A.J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.*, 113(6):3323–3334.

Blauert J. (1997). *Spatial hearing: The psychophysics of sound localization*, Chapter 3. Cambridge, MA: MIT Press.

Boets B., Wouters J., van Wieringen A., Ghesquière P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory *Neuropsychologia*, 45:1608–1620.

Brainard D.H. (1997). The Psychophysics toolbox. *Spatial Vision*, 10(4):433–436.

Brand T., Hohmann V. (2002). An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.*, 112(4):1597–1604.

Bregman A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound,* Chapter 2. Cambridge, MA: MIT Press.

Buckley M.J., Eagleson G.K. (1986). Assessing large sets of rank correlations. *Biometrika*, 73(1):151–157.

Buss E., Hall J.W. III, Grose J.H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear Hear.*, 25(3):242–250.

Buus S., Florentine M., Redden R.B. (1982). The SISI test: A review. *Audiology*, 21:273–293.

Carney, LH. (1994). Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hear. Res*, 76:31–44.

Cedolin L., Delgutte B. (2010). Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. *J. Neurosci.*, 30(38):12712–12724.

Chait M., Eden G., Poeppel D., Simon J.Z., Hill D.F., Flowers D.L. (2007). Delayed detection of tonal targets in background noise in dyslexia. *Brain Lang.*, 102:80–90.

Clarkson M.G., Rogers E.C. (1995). Infants require low-frequency energy to hear the pitch of the missing fundamental. *J. Acoust. Soc. Am.*, 98(1):148–154.

Cohen M.A., Grossberg S., Wyse L.L. (1995). A spectral network model of pitch perception. *J. Acoust. Soc. Am.*, 98:862–879.

Cramer E.M., Huggins W.H.. (1958). Creation of pitch through binaural interaction. *J. Acoust. Soc. Am.*, 30(5):413–417.

Culling J.F. (1999). The existence region of Huggins' pitch. *Hearing Research*, 127:143–148.

Culling J.F., Summerfield A.Q., Marshall D.H. (1998). Dichotic pitches as illusions of binaural unmasking - I - Huggins' pitch and the binaural edge pitch. *J. Acoust. Soc. Am.*, 103(6):3509–3526.

Daneman M., Carpenter P.A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Be.*, 19:450–466.

Davis H., Silverman S.R., McAuliffe D.R. (1951). Some observations on pitch and frequency. *J. Acoust. Soc. Am.*, 23(1):40–42.

de Boer E. (1956a). *On the "residue" in hearing,* pp. 12–89. Ac. Thesis, Univ. Amsterdam.

de Boer E. (1956b). Pitch of inharmonic signals. *Nature*, 178:535–536.

de Boer E. (1976). On the "residue" and auditory pitch perception. *Handbook of sensory physiology*, Vol. V, Pt. 3. Berlin: Springer.

de Cheveigné A. (1998). Cancellation model of pitch perception. *J. Acoust. Soc. Am.*, 103(3):1261–1271.

de Cheveigné A. (2005). Pitch perception models. *Pitch – Neural Coding and Perception,* Chapter 6. New York: Springer Handbook of Auditory Research.

Demany L., Semal C. (1989). Detection thresholds for sinusoidal frequency modulation. *J. Acoust. Soc. Am.*, 85(3):1295–1301.

Depessemier P., Andries C. (2009) *Gletschr, test voor gevorderd lezen et schrijven.* Antwerpen: Garant.

Dix M.R., Hallpike C.S., Hood J.D. (1948). Observations upon the loudness recruitment phenomenon, with especial reference to the differential diagnosis of disorders of the internal ear and VIII nerve. *Proc. R. Soc. Med.*, 41:516–526.

Domnitz R. (1973). The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. *J. Acoust. Soc. Am.*, 53(6):1549–1552.

Dougherty R.F., Cynader M.S., Bjornson B.H., Edgell D., Giaschi D.E. (1998) Dichotic pitch: A new stimulus distinguishes normal and dyslexic auditory function. *NeuroReport*, 9(13):3001–3005.

Dreyer A., Delgutte B. (2006). Phase-locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization. *J. Neurophysiol.*, 96:2327–2341.

DSL (Det Danske Sprog- og Litteraturselskab). (2002). *Korpus 2000 – Dansk sprog omkring årtusindskiftet.*

Durlach N.I. (1960). Note on the equalization-cancellation theory of binaural masking level differences. *J. Acoust. Soc. Am.*, 32:1075–1076.

Edgeworth F.Y. (1908). On the probable errors of frequency-constants. *J. R. Stat. Soc.*, 71:499–512.

Edwards V.T., Giaschi D.E., Dougherty R.F., Edgell D., Bjornson B.H., Lyons C., Douglas R.M. (2004). Psychophysical indexes of temporal processing abnormalities in children with developmental dyslexia. *Dev. Neuropsychol.*, 25(3):321–354.

Feldman A.S. (1977). Diagnostic application and interpretation of tympanometry and the acoustic reflex. *Audiology*, 16:294–306.

Fisher L.D., van Belle, G. (1993). *Biostatistics: A methodology for the health sciences*. Hoboken, NJ: Wiley.

Fletcher H. (1940). Auditory patterns. *Rev. Mod. Phys.*, 12:47–66.

Florentine M., Buus S. (1984). Temporal gap detection in sensorineural and simulated hearing impairments. *J. Speech. Hear. Res.*, 27(3):449–455.

Foo C., Rudner M., Rönnberg J., Lunner T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J. Am. Acad. Audiol.*, 18:618–631.

Fourier J.B.J. (1822). *Traité analytique de la chaleur*. Paris: Didot.

Francart T., van Wieringen A., Wouters J. (2008) APEX 3: A multi-purpose test platform for auditory psychophysical experiments. *J. Neurosci. Meth.*, 172:283–293.

France S.J., Rosner B.S., Hansen P.C., Calvin C., Talcott J.B., Richardson A.J., Stein J.F. (2002). Auditory frequency discrimination in adult developmental dyslexics. *Percept. Psychophys.*, 64(2):169–179.

Gabriel K.J., Koehnke J., Colburn H.S. (1992). Frequency dependence of binaural performance in listeners with impaired binaural hearing. *J. Acoust. Soc. Am.*, 91(1):336–347.

Galilei G. (1638). *Mathematical discourses concerning two new sciences relating to mechanics and local motion, in four dialogues*. Translated by Weston, London: Hooke.

Gardner T.J., Magnasco M.O. (2006). Sparse time-frequency representations. *P. Natl. Acad. Sci. USA*, 103(16):6094–6099.

Gardner B., Martin K. (1994). HRTF measurements of a KEMAR dummy-head microphone. *Technical Report #280.* Cambridge, MA: MIT Media Lab Machine Perceptual Computing.

Gelfand S.A. (2001a). Acoustic immittance assessment. *Essentials of audiology, 2nd ed.,* Chapter 7. Thieme, New York.

Gelfand S.A. (2001b). Behavioral tests for diagnostic assessment. *Essentials of audiology, 2nd ed.,* Chapter 10. Thieme, New York.

Gilbert G., Lorenzi C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *J. Acoust. Soc. Am.*, 119(4):2438–2444.

Glasberg B.R., Moore B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.*, 47:103–138.

Goldstein J.L. (1967). Auditory nonlinearity. *J. Acoust. Soc. Am.*, 41:676–699.

Goldstein J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.*, 109:701–712.

Goverts S.T., Houtgast T. (2010). The binaural intelligibility level difference in hearing-impaired listeners: The role of supra-threshold deficits. *J. Acoust. Soc. Am.*, 127(5):3073–3084.

Grant K.W. (1987). Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners. *J. Speech Hear. Res.*, 30:558–563.

Green D.M., Onsan Z.A., Forrest T.G. (1987). Frequency effects in profile analysis and detecting complex spectral changes. *J. Acoust. Soc. Am.*, 81(3):692–699.

Grubbs F.E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11:1–21.

Hacker M.J., Ratcliff R. (1979). A revised table of d' for M-alternative forced choice. *Percept. Psychophys*, 26(2):168–170.

Hall J.L. (1972). Auditory distortion products f_2-f_1 and $2f_1$-f_2. *J. Acoust. Soc. Am.*, 122(2):1055–1068.

Hall J.W., Buss E., Grose J.H. (2003). Modulation rate discrimination for unresolved components: Temporal cues related to fine structure and envelope. *J. Acoust. Soc. Am.*, 87(1):304–310.

Hämäläinen J.A., Leppänen P.H.T., Guttorm T.K., Lyytinen H. (2008). Event-related potentials to pitch and rise time change in children with reading disabilities and typically reading children. *Clin. Neurophysiol.*, 119:100–115.

Hari R., Renvall H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn. Sci.*, 5(12):525–532.

Hari R., Sääskilahti A., Helenius P., Uutela K. (1999). Non-impaired auditory phase locking in dyslexic adults. *NeuroReport*, 10:2347–2348.

Hartmann W.M. (1984a). A search for central lateral inhibition. *J. Acoust. Soc. Am.*, 75(2):528–535.

Hartmann W.M. (1984b). Binaural coherence edge pitch. *J. Acoust. Soc. Am.*, 75:S22.

Hartmann W.M. (1993). On the origin of the enlarged melodic octave. *J. Acoust. Soc. Am.*, 93(6):3400–3409.

Hartmann W.M., McMillon, C.D. (2001). Binaural coherence edge pitch. *J. Acoust. Soc. Am.*, 109(1):294–305.

Hartmann W.M., Zhang P.X. (2003). Binaural models and the strength of dichotic pitches. *J. Acoust. Soc. Am.*, 114(6, Pt. 1):3317–3326.

Heinz M.G., Colburn H.S., Carney L.H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Comput.*, 13:2273–2316.

Heinz M.G., Swaminathan J. (2009). Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech. *J. Assoc. Res. Otolaryngol.*, 10:407–423.

Hershkowitz R.M., Durlach N.I. (1969). Interaural time and amplitude jnds for a 500-Hz tone. *J. Acoust. Soc. Am.*, 46(6, Pt. 2):1464–1467.

Hill N.I., Bailey P.J., Griffiths Y.M., Snowling M.J. (1999). Frequency acuity and binaural masking release in dyslexic listeners. *J. Acoust. Soc. Am.*, 106(6):L53–L58.

Hood J.D. (1969). Basic audiological requirements in neuro-otology. *J. Laryngol. Otol.*, 83:695–711.

Hopkins K., Moore B.C.J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J. Acoust. Soc. Am.*, 122(2):1055–1068.

Hopkins K., Moore B.C.J., Stone M.A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J. Acoust. Soc. Am.*, 123(2):1140–1153.

Houtsma A.J.M., Goldstein J.L. (1972). The central origin of the pitch of pure tones: Evidence from musical interval recognition. *J. Acoust. Soc. Am.*, 51:520–529.

Houtsma A.J.M., Smurzynski J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.*, 87:304–310.

ISO 16832:2006(E). (2006). *Acoustics – Loudness scaling by means of categories*, Annex A. Geneva: International Organization for Standardization.

Ives D.T., Patterson R.D. (2008). Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *J. Acoust. Soc. Am.*, 123(5):2670–2679.

Jerger J. (1970). Clinical experience with impedance audiometry. *Archiv. Otolaryngol.*, 92(4):311–324.

Jerger J., Shedd J.L., Harford E. (1959). On the detection of extremely small changes in sound intensity. *Archiv. Otolaryngol.*, 69(2):200–211.

Johansson M.S.K., Arlinger S.D. (2002). Binaural masking level difference for speech signals in noise. *Int. J. Audiol.*, 41:279–284.

Johnson D.H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.*, 68:1115–1122.

Joris P.X., Yin T.C.T. (1992). Responses to amplitude-modulated tones in the auditory nerve of the cat. *J. Acoust. Soc. Am.*, 91(1):215–232.

Julesz B. (1971). *Foundations of cyclopean perception.* Chicago, IL: University of Chicago Press.

Kaernbach C. (1991). Simple adaptive testing with the weighted up-down method. *Percept. Psychophys.*, 49:227–229.

Kaernbach C., Bering C. (2001). Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J. Acoust. Soc. Am.*, 110(2):1039–1048.

King K., Stephens D. (1992). Auditory and psychological factors in auditory disability with normal hearing. *Scand. Audiol.*, 21(2):109–114.

Klein M.A., Hartmann W.M. (1981). Binaural edge pitch. *J. Acoust. Soc. Am.*, 70(1):51–61.

Kohlrausch A., Fassel R., van der Heijden M., Kortekaas R., van de Par S., Oxenham A.J., Püschel D. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acustica*, 83:659–669.

Köppl C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the Barn Owl, Tyto. *J. Neurosci.*, 17:3312–3321.

Lacher-Fougère S., Demany L. (1998). Modulation detection by normal and hearing-impaired listeners. *Audiology*, 37:109–121.

Lacher-Fougère S., Demany L. (2005). Consequences of cochlear damage for the detection of interaural phase differences. *J. Acoust. Soc. Am.*, 118(4):2519–2526.

Larsen E., Cedolin L., Delgutte B. (2008). Pitch representations in the auditory nerve: Two concurrent complex tones. *J. Neurophysiol.*, 100:1301–1319.

Levitt H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 49:467–477.

Licklider J.C.R. (1951). A duplex theory of pitch perception. *Experientia*, 7(4):128–134.

Licklider J.C.R. (1954). "Periodicity" pitch and "place" pitch. *J. Acoust. Soc. Am.*, 26(5):945–945.

Licklider J.C.R. (1956). Auditory frequency analysis. *Information theory*, 253–268. Cherry ed. Butterworth: London.

Loeb G.E., White M.W., Merzenich M.M. (1983). Spatial cross-correlation: A proposed mechanism for acoustic pitch perception. *Biol. Cybernetics*, 47:149–163.

Lorenzi C., Gilbert G., Carn H., Garnier S., Moore B.C.J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *P. Natl. Acad. Sci. USA*, 103(49):18866–18869.

Lunner T. (2003). Cognitive function in relation to hearing aid use. *Int. J. Audiol.*, 42:S49–S58.

Lyon G.R., Shaywitz S.E., Shaywitz B.A. (2003). A definition of dyslexia. *Ann. Dyslexia*, 53:1–14.

Mathes R.C., Miller R.L. (1947). Phase effects in monaural perception. *J. Acoust. Soc. Am.*, 19(5):780–797.

McAnally K.I., Stein J.F. (1996). Auditory temporal coding in dyslexia. *Proc. R. Soc. Lond. B*, 263:961–965.

McDermott J.H., Oxenham A.J. (2008). Spectral completion of partially masked sounds. *P. Natl. Acad. Sci. USA*, 105(15):5939–5944.

McLachlan G., Peel D. (2000). *Finite Mixture Models,* Chapter 2. Hoboken, NJ: Wiley.

Meddis R., Hewitt M.J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *J. Acoust. Soc. Am.*, 89(6):2866–2882.

Meddis R., Hewitt M.J., Shackleton T.M. (1990). Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse. *J. Acoust. Soc. Am.*, 87:1813–1816.

Mersenne M. (1636). *Harmonie universelle*. Paris: Cramoisy.

Merzenich M.M., Knight P.L., Roth G.L. (1975). Representation of cochlea within primary auditory cortex in the cat. *J. Neurophysiol.*, 38:231–249.

Micheyl C., Dai H., Oxenham A.J. (2010a). On the possible influence of spectral- and temporal-envelope cues in tests of sensitivity to temporal fine structure. *J. Acoust. Soc. Am.*, 127(3, Pt. 2):1809–1810.

Micheyl C., Divis K., Wrobleski D.M., Oxenham A.J. (2010b). Does fundamental-frequency discrimination measure virtual pitch discrimination? *J. Acoust. Soc. Am.*, 128(4):1930–1942.

Montgomery C.R., Clarkson M.G. (1997). Infants' pitch perception: Masking by low- and high-frequency noises. *J. Acoust. Soc. Am.*, 102(6):3665–3672.

Moore B.C.J. (1973). Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.*, 54(3):610–619.

Moore B.C.J. (2003). *An Introduction to the Psychology of Hearing.* 5th ed. London: Academic Press.

Moore B.C.J. (2007). *Cochlear hearing loss – Physiological, psychological and technical issues*, 2nd ed., Chapter 7. Hoboken, NJ: Wiley.

Moore B.C.J., Glasberg B.R., Baer T. (1997). A model for the prediction of thresholds, loudness and partial loudness. *J. Audio Eng. Soc.*, 45:224–240.

Moore B.C.J., Glasberg B.R., Jepsen M.L. (2009a). Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones. *J. Acoust. Soc. Am.*, 125(5):3194–3204.

Moore B.C.J., Glasberg B.R., Low K.E., Cope T., Cope W. (2006). Effects of level and frequency on the audibility of partials in inharmonic complex tones. *J. Acoust. Soc. Am.*, 120(2):934–944.

Moore B.C.J., Glasberg B.R., Proctor G.M. (1992). Accuracy of pitch matching for pure tones and for complex tones with overlapping or nonoverlapping harmonics. *J. Acoust. Soc. Am.*, 91(6):3443–3450.

Moore B.C.J., Hopkins K., Cuthbertson S. (2009b). Discrimination of complex tones with unresolved components using temporal fine structure information. *J. Acoust. Soc. Am.*, 125(5):3214–3222.

Moore B.C.J., Ohgushi K. (1993). Audibility of partials in inharmonic complex tones. *J. Acoust. Soc. Am.*, 93(1):452–461.

Moore B.C.J., Peters R.W., Glasberg B.R. (1990). Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, 88(1):132–140.

Moore B.C.J., Rosen S.M. (1979). Tune recognition with reduced pitch and interval information. *Q. J. Exp. Psychol.*, 31(2):229–240.

Moore B.C.J., Sęk A. (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *J. Acoust. Soc. Am.*, 100(4, Pt. 1):2320–2331.

Moore B.C.J., Sęk A. (2009a). Sensitivity of the human auditory system to temporal fine structure at high frequencies. *J. Acoust. Soc. Am.*, 125(5):3186–3193.

Moore B.C.J., Sęk A. (2009b). Development of a fast method for determining sensitivity to temporal fine structure. *Int. J. Audiol.*, 48:161–171.

Moore B.C.J., Skrodzka E. (2002). Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. *J. Acoust. Soc. Am.*, 111(1, Pt. 1):327–335.

Moore G.A., Moore B.C.J. (2003). Perception of the low pitch of frequency-shifted complexes. *J. Acoust. Soc. Am.*, 112(2):977–985.

Nitschmann M., Verhey J.L., Kollmeier B. (2010). Monaural and binaural frequency selectivity in hearing-impaired subjects. *Int. J. Audiol.*, 100(49):357–367.

Ohm G.S. (1843). On the definition of a tone with the associated theory of the siren and similar sound producing devices. *Poggendorf's Annalen der Physik und Chemie*, 59:497ff.

Oxenham A.J., Bernstein J.G.W., Penagos H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *P. Natl. Acad. Sci. USA*, 101(5):1421–1425.

Oxenham A.J., Micheyl C., Keebler M.V. (2009). Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *J. Acoust. Soc. Am.*, 125(4):2189–2199.

Oxenham A.J., Micheyl C., Keebler M.V., Loper A., Santurette S. (2011). Pitch perception beyond the traditional existence region of pitch. *P. Natl. Acad. Sci. USA*, 108(18):7629-7634.

Palmer A.R., Russell I.J. (1986). Phase locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. *Hear. Res.*, 24:1–15.

Patterson R.D. (1973). The effects of relative phase and the number of components on residue pitch. *J. Acoust. Soc. Am.*, 53(6):1565–1572.

Patterson R.D. (1976). Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.*, 59:640–654.

Patterson R.D., Allerhand M.H., Giguère C. (1995). Time-domain modeling of peripheral auditory processing: A model architecture and a software platform. *J. Acoust. Soc. Am.*, 98:1890–1894.

Patterson R.D., Moore B.C.J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. *Frequency Selectivity in Hearing.* London: Academic Press.

Patterson R.D., Nimmo-Smith I. (1980). Off-frequency listening and auditory-filter asymmetry. *J. Acoust. Soc. Am.*, 67:229–245.

Patterson R.D., Nimmo-Smith I., Weber D.L., Milroy R. (1982). The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.*, 72:1788–1803.

Peissig J., Kollmeier B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J. Acoust. Soc. Am.*, 101(3):1660–1670.

Pelli D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4):437–442.

Penner M.J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *J. Acoust. Soc. Am.*, 61:552–557.

Plack C.J., Carlyon R.P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J. Acoust. Soc. Am.*, 98:1355–1364.

Plack C.J., Oxenham A.J. (2005a). Overview: The present and future of pitch. *Pitch – Neural Coding and Perception*, Chapter 1. New York: Springer Handbook of Auditory Research.

Plack C.J., Oxenham A.J. (2005b). The psychophysics of pitch. *Pitch – Neural Coding and Perception*, Chapter 2. New York: Springer Handbook of Auditory Research.

Plomp R. (1964). The rate of decay of auditory sensation. *J. Acoust. Soc. Am.*, 36:277–282.

Plomp R. (1967). Pitch of complex tones. *J. Acoust. Soc. Am.*, 41(6):1526–1533.

Plomp R., Mimpen A.M. (1968). The ear as a frequency analyzer. II. *J. Acoust. Soc. Am.*, 43(4):764–767.

Pollack I. (1969). Periodicity pitch for white noise – Fact or artifact? *J. Acoust. Soc. Am.*, 45:237–238.

Priede V.M., Coles R.R.A. (1974). Interpretation of loudness recruitment tests. *J. Laryngol. Otol.*, 88(2):641–662.

Probst R., Lonsbury-Martin B.L., Martin G.K. (1991). A review of otoacoustic emissions. *J. Acoust. Soc. Am.*, 89(5):2027–2067.

Raatgever J., Bilsen F.A. (1986). A central spectrum theory of binaural processing. Evidence from dichotic pitch. *J. Acoust. Soc. Am.*, 80(2):429–441.

Ramus F. (2003). Developmental dyslexia: Specific phonological deficit or general sensori-motor dysfunction? *Curr. Opin. Neurobiol.*, 13:212–218.

Ritsma R.J. (1962). Existence region of the tonal residue. I. *J. Acoust. Soc. Am.*, 34:1224–1229.

Ritsma R.J. (1967). Frequencies Dominant in the Perception of the Pitch of Complex Sounds. *J. Acoust. Soc. Am.*, 42(1):191–198.

Ritsma R.J., Engel F.L. (1964). Pitch of frequency-modulated signals. *J. Acoust. Soc. Am.*, 36(9):1637–1644.

Roberts B., Bregman A.S. (1991). Effects of the pattern of spectral spacing on the perceptual fusion of harmonics. *J. Acoust. Soc. Am.*, 90(6):3050–3060.

Rose J.E., Brugge J.F., Anderson D.J., Hind J.E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of squirrel monkey. *J. Neurophysiol.*, 30:769–793.

Rose J.E., Galambos R., Hughes J.R. (1959). Microelectrode studies of the cochlear nuclei of the cat. *B. Johns Hopkins Hosp.*, 104(5):211–251.

Rosen S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *J. Phonetics*, 31:509–527.

Ross B., Fujioka T., Tremblay K.L., Picton T.W. (2007a). Aging in binaural hearing begins in mid-life: Evidence from cortical auditory-evoked responses to changes in interaural phase. *J. Neurosci.*, 27(42):11172–11178.

Ross B., Tremblay K.L., Picton T. (2007b). Physiological detection of interaural phase differences. *J. Acoust. Soc. Am.*, 121(2):1017–1027.

Ruggero M.A., Rich N.C. (1991). Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane. *J. Neurosci.*, 11:1057–1067.

Rutherford E. (1886). A new theory of hearing. *J. Anat. Physiol.*, 21:166–168.

Salvi R.J., Arehole S. (1985). Gap detection in chinchillas with temporary high-frequency hearing loss. *J. Acoust. Soc. Am.*, 77(3):1173–1177.

Santurette S., Dau T. (2007). Binaural pitch perception in normal-hearing and hearing-impaired listeners. *Hear. Res.*, 223(1-2):29–47.

Santurette S., Dau T. (2011a). The role of temporal fine structure for the low pitch of high-frequency complex tones. *J. Acoust. Soc. Am.*, 129(1):282-292.

Santurette S., Dau T. (2011b). Relating binaural pitch perception to the individual listener's auditory profile. *J. Acoust. Soc. Am.*, submitted.

Santurette S., Poelmans H., Luts H., Ghesquière P., Wouters J., Dau T. (2010). Detection and identification of monaural and binaural pitch contours in dyslexic listeners. *J. Assoc. Res. Otoloaryngol.*, 125(11):515–524.

Saunders G.H., Haggard M.P. (1989). The clinical assessment of obscure auditory dysfunction - 1. Auditory and psychological factors. *Ear Hear.*, 10:200–208.

Schouten J.F. (1938). The perception of subjective tones. *Proc. Kon. Acad. Wetensch.*, 41:1086–1094.

Schouten J.F. (1940). The residue, a new component in subjective sound analysis. *Proc. Kon. Acad. Wetensch.*, 43:356–365.

Schouten J.F., Ritsma R.J., Cardozo B.L. (1962). Pitch of the residue. *J. Acoust. Soc. Am.*, 34:1418–1424.

Schroeder M.R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *J. Acoust. Soc. Am.*, 43(4):829–834.

Schweder T., Spjøtvoll E. (1982). Plots of P-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502.

Seebeck A. (1841). Beobachtungen über einige Bedidungen der Entstehung von Tönen. *Ann. Phys. Chem.*, 53:417–436.

Shackleton T.M., Carlyon R.P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.*, 95(6):3529–3540.

Shamma S.A. (1985). Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.*, 78(5):1622–1632.

Shamma S.A. (2001). On the role of space and time in auditory processing. *Trends Cog. Sci.*, 5(8):340–348.

Shamma S.A. (2004). Topographic organization is essential for pitch perception. *P. Natl. Acad. Sci. USA*, 101(5):1114–1115.

Shamma S.A., Klein D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.*, 107(5),Pt.1:2631–2644.

Shamma S.A., Shen N., Gopalaswamy P. (1989). Stereausis: Binaural processing without neural delays. *J. Acoust. Soc. Am.*, 86(3):989–1006.

Shaywitz S.E., Shaywitz B.A., Fletcher J.M., Escobar M.D. (1990). Prevalence of reading disability in boys and girls. Results of the Connecticut longitudinal study. *J. Amer. Med. Assoc.*, 264(8):998–1002.

Slaney M. (1998). Auditory toolbox: A MATLAB toolbox for auditory modeling work, version 2. *Technical Report #1998-010*. Palo Alto, CA: Interval Research Corporation.

Smoorenburg G.F. (1970). Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.*, 48(4, Pt. 2):924–942.

Staffel J.G., Hall J.W., Grose J.H., Pillsbury H.C. (1990). NoSo and NoSπ detection as a function of masker bandwidth in normal-hearing and cochlear-impaired listeners. *J. Acoust. Soc. Am.*, 87(4):1720–1727.

Strelcyk O., Dau T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.*, 125(5):3328–3345.

Terhardt E. (1972). Zur Tonhöhenwahrnehmung von Klängen (Perception of the pitch of complex tones). *Acustica*, 26:173–199.

Terhardt E. (1974). Pitch, consonance and harmony. *J. Acoust. Soc. Am.*, 55:1061–1069.

Thurlow W.R., Small A.M.Jr. (1955). Pitch perception for certain periodic harmonic stimuli. *J. Acoust. Soc. Am.*, 27(1):132–137.

Tyler R.S., Wood E.J., Fernandes M. (1983). Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *J. Acoust. Soc. Am.*, 74(4):1190–1199.

van den Bos K.P., Spelberg H.C.L., Scheepstra A.J.M., De Vries J.R. (1994). *De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling.* Nijmegen: Berkhout.

van de Par S., Kohlrausch A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *J. Acoust. Soc. Am.*, 101(3):1671–1680.

Vandermosten M., Boets B., Luts H., Poelmans H., Golestani N., Wouters J., Ghesquière P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *P. Natl. Acad. Sci. USA*, 107(23):10389–10394.

van Rooij J.C.G.M., Plomp R., Orlebeke J.F. (1989). Auditive and cognitive factors in speech perception by elderly listeners. I: Development of test battery. *J. Acoust. Soc. Am.*, 86(4):1294–1309.

van Rooij J.C.G.M., Plomp R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *J. Acoust. Soc. Am.*, 88(6):2611–2624.

Voeckler H.B. (1966). Toward a unified theory of modulation. Part I: Phase-envelope relationships. *Proc. IEEE*, 54(3):340–355.

von Helmholtz H. (1877). *On the sensations of tone*. English translation A. J. Ellis (1885, 1954). New York: Dover.

Wagener K., Josvassen J.L., Ardenkjær R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *Int. J. Audiol.*, 42:10–17.

Walliser K. (1969). A functional scheme for formation of pitch residue from an acoustical stimulus. *Kybernetik*, 6(2):65–72.

Ward W.D. (1954). Subjective musical pitch. *J. Acoust. Soc. Am.*, 26(3):369–380.

Wechsler D. (1999). *Wechsler adult intelligence scale.* Third ed. NL. London: The Psychological Corporation.

Wever E.G., Bray C.W. (1930). The nature of acoustic response: The relation between sound frequency and frequency of impulses in the auditory nerve. *J. Exp. Psychol.*, 13:373–387.

Wier C.C., Jesteadt W., Green D.M. (1977). Frequency discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.*, 61(1):178–184.

Wightman F.L. (1973). The pattern-transformation model of pitch. *J. Acoust. Soc. Am.*, 54(2):407–416.

Wright B.A., Bowen R.W., Zecker S.G. (2000). Nonlinguistic perceptual deficits associated with reading and language disorders. *Curr. Opin. Neurobiol.*, 10:482–486.

Yost W.A. (1991). Thresholds for segregating a narrow-band from a broadband noise based on interaural phases and level differences. *J. Acoust. Soc. Am.*, 89:838–844.

Zeng F.G., Kong Y.Y., Michalewski H.J., Starr A. (2005). Perceptual consequences of disrupted auditory nerve activity. *J. Neurophysiol.*, 93:3050–3063.

Zeng F.G., Nie K., Liu S., Stickney G., Del Rio E., Kong Y.Y., Chen H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. *J. Acoust. Soc. Am.*, 116(3):1351–1354.

Zhang P.X., Hartmann W.M. (2004). Lateralization of the Huggins pitch. *J. Acoust. Soc. Am.*, 115(abst):2534(4aPP6).

Zhang X., Heinz M.G., Bruce I.C., Carney L.H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.*, 109(2):648–670.

Zwicker E. (1981). Dependence of level and phase of the ($2f_1$-f_2)-cancellation tone on frequency range, frequency difference, level of primaries, and subject. *J. Acoust. Soc. Am.*, 70(5):1277–1288.

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.

The end.

To be continued...

Pitch is an important attribute of hearing that allows us to perceive the musical quality of sounds. Besides music perception, pitch contributes to speech communication, auditory grouping, and perceptual segregation of sound sources. This work investigated several aspects of pitch perception in the human auditory system via psychophysical experiments with normal-hearing and hearing-impaired listeners. The central question of the relative importance of spectral and temporal information for pitch perception was addressed by using artificial stimuli with specific spectrotemporal properties. The effects of hearing impairment on binaural pitch perception were also studied, with an emphasis on the role of specific auditory functions. This work provides insights into the fundamental auditory mechanisms underlying pitch perception and may have implications for future pitch-perception models, as well as strategies for auditory-profile characterization and restoration of accurate pitch perception in impaired hearing.

Centre for Applied Hearing Research
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

ISBN 978-87-92465-90-0