KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK – ESAT
Kasteelpaark Arenberg 10, 3001 Leuven (Heverlee)

In samenwerking met
FACULTEIT GENEESKUNDE
DEPARTEMENT NEUROWETENSCHAPPEN – Lab. Exp. ORL
Kapucijnenvoer 33, 3000 Leuven

WHEN THE DEAF LISTEN TO MUSIC

PITCH PERCEPTION WITH COCHLEAR IMPLANTS

Jury:
Prof. Dr. ir. P. Van Houtte, voorzitter
Prof. Dr. J. Wouters, promotor
Prof. Dr. ir. M. Moonen, promotor
Prof. Dr. ir. D. Van Compernolle
Prof. Dr. ir. S. Van Huffel
Prof. Dr. ir. T. Houtgast (VU, Amsterdam)
Prof. Dr. ir. S. Peeters (UA, Antwerpen)
Dr. R.P. Carlyon (MRC-CBU, Cambridge)
Dr. ir. L. Geurts (Groep T, Leuven)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Johan LANEAU

U.D.C. 616.28-089.84 Mei 2005

© Katholieke Universiteit Leuven
Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm, elektronische of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

D/2005/7515/25

ISBN 90-5682-591-7

Voor Lotje en Engelien

Voorwoord

Bij het begin van dit proefschrift wil ik de mensen bedanken die bijgedragen hebben tot het tot stand komen van deze tekst. Als onderzoeker sta je immers niet alleen, maar ben je omringd door mensen die je steunen, raad en opbouwende kritiek geven, en je oppeppen.

In de eerste plaats gaat mijn oprechte dank uit naar Prof. Jan Wouters. Hij gaf me de kans om een jongensdroom van me te verwezenlijken, namelijk *uitvinder* worden. Toen hij me enkele jaren geleden de kans bood een doctoraat te behalen klonk dit als muziek in mijn oren. Ook al omdat het ging over het klinken van muziek in dove oren. In de voorbije jaren heb ik Jan in de eerste plaats leren appreciëren als een zeer enthousiasmerende persoon. De energie die hij opwekte bij mij, en bij andere collega's, ontsproot waarschijnlijk uit de passie voor onderzoek en voor wetenschap die hem eigen is. Deze passie zorgde ervoor dat je kon binnenstappen in zijn bureau met (schijnbaar) tegenvallende resultaten, en na een goede discussie zou je weer buitenstappen vol inspiratie omdat de eerst onverstaanbare resultaten net een mogelijke ontdekking konden zijn. Maar Jan is ook een aangenaam mens, met soms wat jongensachtige streken die ik met hem deel. Hierdoor beleefden we ook veel plezier de voorbije jaren: aan het onderzoek en aan de ontspanning zoals de duikboot-kano.

Mijn tweede promotor, Prof. Marc Moonen wil ik bedanken voor de bijdragen die hij leverde vanuit zijn eigen perspectief dat complementair was aan het perspectief van de mensen in het Lab. Exp. ORL. Zijn zin voor wiskundige nauwkeurigheid heeft ertoe geleid dat de beschrijving van het onderzoek in dit proefschrift duidelijker is en dat de experimenten makkelijker reproduceerbaar zijn.

Ik wens de leden van mijn begeleidingscommissie, Prof. Sabinne Van Huffel en Prof. Dirk Van Compernolle, te bedanken voor het kritisch nalezen van mijn proefschrift en de waardevolle suggesties en aanvullingen. Ik dank ook de andere juryleden voor hun bereidheid om deel uit te maken van de examencommissie. Ik

dank Prof. Tammo Houtgast om aanwezig te kunnen zijn op de openbare verdediging. Prof. Stef Peeters dank ik voor zijn onafgebroken pionierswerk op het vlak van cochleaire implantaten in Vlaanderen. Het bestaan in Vlaanderen van verschillende onderzoeksgroepen en bedrijven die werken rond cochleaire implantaten is hoogst waarschijnlijk te danken aan zijn inspanningen om het Laura implantaat te ontwikkelen. I wish to thank Dr. Bob Carlyon for a number of very helpful discussions on the collected data and for the critical reading of my early manuscripts. Some of his remarks, based upon his encyclopaedical knowledge of the scientific literature on pitch perception, proved to be very valuable and put me in the right direction for the interpretation of a lot of results. Dr. Luc Geurts wil ik bedanken voor de praktische richtlijnen die hij gaf bij het begin van dit doctoraat. Veel van de studies in dit proefschrift bouwen rechtstreeks verder op zijn ideeën en onderzoek.

Ik wens ook uitdrukkelijk mijn dank te betuigen aan de proefpersonen die deelgenomen hebben aan de vaak moeilijke en eentonige experimenten. Ik ben hen ontzettend dankbaar voor de motivatie die ze toonden voor de vele uren dat ze naar *piepjes* luisterden. Zonder hun medewerking zou dit proefschrift nooit tot stand zijn gekomen. Vandaar mijn grote waardering voor hun onbaatzuchtigheid, en hun bijdrage aan de wetenschap.

Tijdens dit proefschrift heb ik nauw samengewerkt met Cochlear Ltd., en dan vooral met hun zetel CTCE in Mechelen. Van deze mensen wil ik vooral Bas van Dijk, Jan Janssen, Colin Irwin, Brett Swanson en Egwin van Den Heuvel bedanken. De hulpmiddelen en de informatie die zij verschaften waren noodzakelijk voor de verwezenlijking van dit proefschrift. Hierbij wil ik ook Cochlear Ltd. samen met het Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT) bedanken voor de financiering van het onderzoek dat geleid heeft tot dit proefschrift.

Gedurende de voorbije jaren ben ik omringd geweest door zeer aangename collega's die zorgden voor een zeer goede sfeer. Ik ben hen dankbaar voor de steun en hulp die ze me verstrekten. In het bijzonder wil ik Astrid, Bram, Heleen, Jean-Baptiste, Koen, Lies, Lot, Olivier, Stijn, Tim, en Tom bedanken voor de fijne uurtjes in het Lab. Exp. ORL. Ook bedankt aan de mensen in SISTA voor hun hulp bij mathematische of praktische probleempjes. Verder was er nog de uitstekende technische ondersteuning van Dennis en Wilfried. Ook ben ik dankbaar voor de hulp en toelichtingen die ik kreeg van de afdeling Neus, Keel, Oorziekten, Gelaat- en Halschirurgie van het UZ Leuven, het Revalidatiecentrum Gehoor en Spraak en de mensen van de opleiding Logopedie en Audiologie.

Op persoonlijk vlak wil ik vrienden en familie bedanken voor de steun en voor de hulp bij het ontspannen. Een speciaal woord van dank gaat hierbij naar mijn ouders die ik verdenk van de jongensdroom om uitvinder te worden (al dan niet onbewust) in mijn hoofd gestoken te hebben. Zij hebben mij altijd gesteund in de keuzes die ik gemaakt heb, of uitgedrukt in hun eigen woorden: Zij hebben me diepe wortels gegeven en de vleugels om ver te vliegen.

Het laatste dankjewel is een speciale en gaat uit naar Lotje. Zij heeft mij, samen met Engelien, een thuis gegeven waar het mogelijk was om dit werk tot een goed einde

te brengen. Zij hielp mij over tegenslagen in het onderzoek en we genoten samen wanneer er grote vooruitgang was. De vervulling van mijn jongensdroom kreeg een extra, fantastische dimensie dankzij Lotjes aanwezigheid gedurende de hele rit. Ik hoop dat we samen nog vele dromen zullen verwezenlijken. Ik draag dit proefschrift op aan haar en aan Engelien.

Abstract

Cochlear implants (CI) are surgically implanted hearing aids that provide auditory sensations to deaf people through direct electrical stimulation of the auditory nerve. Although relatively good speech understanding can be achieved by implanted subjects, pitch perception by CI subjects is about 50 times worse than observed for normal-hearing (NH) persons. Pitch is, however, important for intonation, music, speech understanding in tonal languages, and for separating multiple simultaneous sound sources. The major goal of this work is to improve pitch perception by CI subjects.

In CI subjects two fundamental mechanisms are used for pitch perception: place pitch and temporal pitch. Our results show that place pitch is correlated to the sound's brightness because place pitch sensation is related to the centroid of the excitation pattern along the cochlea. The slopes of the excitation pattern determine place pitch sensitivity. Our results also show that the effectiveness of temporal pitch improves as the modulation depth increases and the different channels are modulated synchronously. We also show that both temporal and place pitch cues, although independent sensations, can be integrated and projected on a "combined pitch"-axis used for fundamental frequency (F0) discrimination. (The pitch height NH persons experience corresponds (most often) to the F0 of the sound.) The contribution of place pitch cues for F0-discrimination is, however, hindered by the electrical current spread along the cochlea producing shallow slopes of the excitation pattern and by the interference of the pitch cues with the spectral envelope. The temporal pitch mechanism provides the most effective cues for F0-discrimination but it is limited by more central neural processing.

Although we show that the effectiveness of either temporal or place pitch cues can be improved for some particular filter banks, we also show that the simultaneous effective presentation of both temporal and place pitch cues for F0-discrimination is not possible with current sound-processing schemes for CI. Therefore, we developed in this thesis a new sound-processing scheme designed to optimize pitch sensation:

F0mod. In this scheme the pitch of the processed sound is explicitly encoded in the electrical stimulation pattern delivered to the subjects. With the new F0mod scheme we found a 3-fold improvement for pitch discrimination with respect to the current sound-processing scheme and that CI subjects recognize more melodies with the F0mod scheme.

The investigation of the pitch mechanisms in CI subjects also enabled us to build an acoustic model of CI pitch sensation. In this way NH persons can have an idea of the pitch quality CI subjects' experience.

Korte inhoud

Cochleaire implantaten (CI) zijn een type hoorapparaat dat toelaat om dove mensen opnieuw te laten horen door middel van rechtstreekse elektrische stimulatie van de gehoorzenuw. Hoewel een CI toelaat om spraak te verstaan in een stille omgeving, is de toonhoogtewaarneming bij geïmplanteerde personen ongeveer 50 maal slechter dan bij normaalhorende (NH) personen. Toonhoogte is echter belangrijk voor intonatie, muziek, spraakverstaan bij tonale talen, en voor het onderscheiden van meerdere gelijktijdige geluidsbronnen. Het hoofddoel van dit proefschrift is het verbeteren van de toonhoogtewaarneming door CI patiënten.

CI patiënten gebruiken twee basismechanismen voor toonhoogtewaarneming: plaats-afhankelijke toonhoogte en temporele toonhoogte. Onze resultaten tonen aan dat de plaatsafhankelijke toonhoogte (of plaats-toonhoogte) bepaald wordt door het zwaartepunt van het elektrisch stimulatiepatroon langsheen de cochlea. De steilheid van de flanken van het stimulatiepatroon bepaalt de gevoeligheid van de patiënt voor veranderingen in plaats-toonhoogte. We tonen ook aan dat de efficiëntie van het temporele toonhoogtemechanisme vergroot bij grotere modulatiediepte en wanneer de verschillende kanalen synchroon gemoduleerd zijn. Onze resultaten tonen verder aan dat de twee toonhoogtemechanismen kunnen gecombineerd kunnen worden in een globale toonhoogtewaarneming die gebruikt wordt voor de discriminatie van de fundamentele frequentie (F0). De toonhoogte die NH personen waarnemen komt overeen met de fundamentele frequentie (F0) van het geluid. De bijdrage van het "plaats-toonhoogte"-mechanisme is echter zwak omwille van de interactie van het mechanisme met de omhullende van het spectrum en omwille van de vrij vlakke flanken van het excitatiepatroon door het uitsmeren van de elektrische stroom in de cochlea. Het temporele toonhoogtemechanisme draagt het meest efficiënt bij tot F0-discriminatie maar is beperkt door de meer centrale neurale verwerking.

Hoewel we aantonen dat de doeltreffendheid voor F0-discriminatie van ofwel de plaats-toonhoogte ofwel de temporele toonhoogte kan verbeterd worden d.m.v. specifieke filterbanken, tonen we ook aan dat het niet mogelijk is met de huidige

signaalverwerkingsschemas deze beide mechanismen optimaal te benutten. Daarom is in dit proefschrift een nieuw geluidsverwerkingsschema ontwikkeld dat ontworpen is om de toonhoogtewaarneming te optimaliseren. Dankzij dit nieuwe schema kunnen CI patiënten driemaal kleinere toonhoogteverschillen onderscheiden en herkennen ze meer melodieën.

Glossary

Mathematical symbols

a, b, c, d	Scalars
A	Amplitude of the envelope
$A_n(t)$	Amplitude of the envelope of channel n at time instant t
$\overline{A_n}$	Average amplitude of channel envelope (averaged over time)
$a_c(\tau, t)$	Auto-correlation function as a function of lag and for the window at time instant t
α	Constant in the expression to convert frequency to cochlear position [Greenwood, 1990]
B_w	Bandwidth
$C_n(t)$	Compressed amplitude of the envelope of channel n at time instant t
$\mathrm{C_n}$	C-level or comfortable loudness level for channel n
C^{acoust}	Centroid of the acoustically elicited neural excitation pattern
C^{elec}	Centroid of the electrically elicited neural excitation pattern
cl_n	Amplitude of the electrical stimulation for channel n expressed in current levels
d'	d-prime value – refers to the discriminablity of two stimuli
d'_{place}	d-prime value for the place pitch cues
d'_{temp}	d-prime value for the temporal pitch cues
$d'_{combined}$	d-prime value for the total of the combined place and temporal pitch cues

Δf	Difference in frequency
Δf_c	Difference in center frequency
Δf_m	Difference in modulation frequency
$\Delta place$	Difference in place pitch estimate
$\Delta spectral$	Difference in spectral pitch estimate
e	Electrode number
f	Frequency
$F(f)$	Frequency response of a filter as a function of frequency
F_c	Center frequency of a filter or channel
F_n	Filter of channel n
$F_n(t)$	Amplitude output of the filter of channel n at time instant t
F_r	Resonance frequency
F_s	Sampling frequency
F_{update}	Update rate or analysis rate of a block based signal-processing system
$F0$	Fundamental frequency
$F0_{est}$	Estimated fundamental frequency
$F1$	(First) Formant frequency
$\Phi_n(q)$	Spectrum of the filter bank outputs for channel n
g_n	Gain of channel n
$H(z)$	Frequency response of a filter in the digital domain
i	Index
$I(t)$	Electrical current as a function of time t
k	Time index for a discrete time signal
λ	Space constant
L	Length of the analysis window
$M^{acoust}(x)$	Magnitude of the acoustically elicited neural excitation pattern as a function of cochlear location
$M^{elec}(x)$	Magnitude of the electrically elicited neural excitation pattern as a function of cochlear location
n	Channel number
N	Total number of channels
p	Proportion (of correctly identified trials)
$r_n[k]$	Rectified signal
$S(f)$	Frequency spectrum as a function of frequency (most often obtained through FFT)
$S(q,t)$	Frequency spectrum of the time frame at time instant t as a function of frequency
t	Time instant

T_n	T-level or threshold of hearing level for channel n
T_s	Sampling period
$T_{stimulus}$	Total length of a stimulus
τ	Lag in the auto-correlation function
$\tau_{\max}$	Lag in the auto-correlation function at which the amplitude of the auto-correlation function is maximal
q	Index of the frequency bin in a FFT spectrum
$w[k]$	Discrete-time window
$W(q)$	Frequency spectrum of a window
x	Cochlear location expressed from apex to base
$x(f)$	Function converting frequency to cochlear location based upon Greenwood's formula [Greenwood, 1990]
$x[k]$	The discrete-time input signal
$y_n[k]$	The output of a discrete-time filter
Z	The inverse cumulative normal Gaussian distribution

Acronyms and abbreviations

2AFC	A two alternative forced-choice experimental design
2I-2AFC	A two interval, two alternative forced-choice experimental design
4I-4AFC	A four interval, four alternative forced-choice design
ACE	Advanced Combination Encoders; sound-processing scheme for cochlear implants of Cochlear Ltd.
AGC	Automatic gain control
ANOVA	Analysis of Variance
APEX	Application for Psycho-Electrical eXperiments
BTE	Behind-The-Ear (hearing aid)
CA	Compressed Analog
CI	Cochlear implant
CIS	Continuous Interleaved Sampler ; sound-processing scheme for cochlear implants
CISIM	Cochlear Implant Simulating
C-level	Comfortable loudness level
CVC	Consonant-Vowel-Consonant words
dB HL	decibel Hearing Level, expressed relative to the hearing threshold of normal-hearing persons
dB SPL	decibel Sound Pressure Level, expressed relative to a reference pressure of 20 μPa
DC	Direct current
DSP	Digital signal processor

e.g.	Exempli gratia (Latin: for example)
ERB	Equivalent Rectangular Bandwidth
et al.	Et alia (Latin : and others)
F0mod	F0 modulated sound processing scheme
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
fMRI	functional Magnetic Resonance Imaging
i.e.	Id est (Latin: that is)
IIR	Infinite Impulse Response
JND	Just Noticeable Difference
LSD	Least Significant Difference
MIDI	Musical Instrument Digital Interface
MP	MonoPolar
NH	Normal-Hearing
BP	BiPolar
PCM	Pulse Code Modulated
PET	Positron Emission Tomography
pps	Pulses per second
RF	Radio Frequency
RMS	Root-Mean-Square
SFS	Single-Formant Stimuli
T-level	Threshold of hearing level

Contents

Voorwoord 5

Abstract 9

Korte inhoud 11

Glossary 13

Mathematical symbols 13

Acronyms and abbreviations 15

Contents 17

Als doven naar muziek luisteren - Toonhoogtewaarneming met een cochleair implantaat 23

1. Inleiding 23
1.1. Doof 23
1.2. Hoe werkt ons gehoor? 24
1.3. Cochleaire implantaten 26
1.4. Toonhoogte 28
1.5. Neurale mechanismen voor toonhoogte 29
1.6. Toonhoogte en muziek bij CI gebruikers 31
1.7. Overzicht van dit proefschrift 32

2. Het onderzoeksplatform: APEX 33

3. Plaatstoonhoogte met stimulatie van meerdere elektroden 34

4. De combinatie van plaatstoonhoogte en temporele toonhoogte 35

5. Het effect van de filterbank op het onderscheiden van F0 36

6. Simulatie van de toonhoogte van een CI geluid 38

7. Verbeterde muziek door nieuwe F0mod geluidsverwerking 39

8. Besluiten 41

Chapter 1 Introduction to cochlear implants and pitch perception 43

1.1 Cochlear implants 43
1.1.1 Hearing loss 43
1.1.2 How does hearing work? 44
1.1.3 History of cochlear implants 47
1.1.4 The cochlear implant device 47
1.1.5 Sound processing in cochlear implants 50
1.1.6 Outcomes of cochlear implantation 54

1.2 Pitch perception in normal-hearing persons 55
1.2.1 What is pitch? 55
1.2.2 Pure tones and harmonic signals 57
1.2.3 Pitch mechanisms 59
1.2.4 Measurement techniques for pitch 62

1.3 Music and pitch perception in cochlear implant subjects 63
1.3.1 Music perception with a cochlear implant 63
1.3.2 Mechanisms of pitch perception in cochlear implant recipients 64
1.3.3 Relations between pitch mechanisms in NH and in CI 67

1.4 Outline of thesis 68
1.4.1 Major research questions 68
1.4.2 Chapter by chapter overview 70

1.5 Conclusion 71

Chapter 2 The research platform: APEX 73

2.1 Introduction 74

2.2 Overview of the software 75
2.2.1 Stimuli 75
2.2.2 Experimental procedure 76
2.2.3 Running the experiment 76
2.2.4 Analysis of results 77

2.3 Internal structure 77

2.4 Animated mode 78

2.5 Outcomes 82

2.6 Summary 82

Chapter 3 Multi-channel place pitch sensitivity in cochlear implant subjects 85

3.1 Introduction 87

3.2 Methods *88*
3.2.1 Subjects 88
3.2.2 Stimuli 89
3.2.3 Loudness balancing 90
3.2.4 Psychophysical procedure 92
3.2.5 Training 93
3.2.6 Data analysis 93
3.3 Results *94*
3.4 Discussion *96*
3.5 Conclusions *97*

Chapter 4 The combination of consistent and inconsistent place and temporal pitch cues in a pitch ranking task 99
4.1 Introduction *100*
4.2 Methods *101*
4.2.1 Subjects 101
4.2.2 Stimuli 102
4.2.3 Psychophysical procedure 103
4.2.4 Data analysis 104
4.3 Results *104*
4.4 Discussion *107*
4.5 Conclusion *110*

Chapter 5 The effect of filter banks on fundamental frequency discrimination by cochlear implant subjects 111
5.1 Introduction *112*
5.2 Methods *114*
5.2.1 Subjects 114
5.2.2 Stimuli 114
5.2.3 Sound processing 115
5.2.4 Psychophysical procedure 125
5.3 Results *126*
5.3.1 Without temporal pitch cues 128
5.3.2 Including temporal pitch cues 128
5.4 Discussion *130*
5.4.1 Without temporal pitch cues 130
5.4.2 Including temporal pitch cues 136
5.5 Conclusions *139*

Chapter 6 Acoustic models for pitch perception by cochlear implant subjects 141

6.1 Introduction *142*

6.2 Sound processing scheme *144*
6.2.1 Analysis-resynthesis vocoder 144
6.2.2 CISIM resynthesis filter 146

6.3 Experiment 1A: Simulated place pitch discrimination *147*
6.3.1 Methods 148
6.3.2 Results 149
6.3.3 Discussion 150

6.4 Experiment 1B: Simulated temporal pitch discrimination *151*
6.4.1 Methods 151
6.4.2 Results 152
6.4.3 Discussion 153

6.5 Experiment 2: F0-discrimination using noise-band vocoders *154*
6.5.1 Methods 154
6.5.2 Results 158
6.5.3 Discussion 163

6.6 General discussion *170*

6.7 Conclusions *172*

Chapter 7 Improved music perception with F0mod sound processing strategy 175

7.1 Introduction *176*

7.2 Sound processing scheme *178*
7.2.1 Rationale of the F0mod scheme 178
7.2.2 Implementation of the F0mod scheme 179
7.2.3 Comparison strategies 184

7.3 Experiment 1: F0-discrimination of single-formant stimuli *185*
7.3.1 Methods 185
7.3.2 Results 187
7.3.3 Discussion 188

7.4 Experiment 2: F0-discrimination for musical notes *189*
7.4.1 Methods 189
7.4.2 Results 190
7.4.3 Discussion 191

7.5 Experiment 3: Melody recognition *193*
7.5.1 Methods 193
7.5.2 Results 194
7.5.3 Discussion 196

7.6 Experiment 4: Musical interval identification *197*
7.6.1 Methods 197
7.6.2 Results 198
7.6.3 Discussion 199

7.7 General discussion *201*
7.7.1 Implications for CI sound processors 201
7.7.2 Mechanisms of pitch perception 202

7.8 Conclusion *202*

7.9 Appendix: The relation between spectral and place pitch *204*

Chapter 8 Conclusions and future research **205**

8.1 The mechanisms for pitch perception in CI subjects *205*
8.1.1 Place pitch 206
8.1.2 Temporal pitch 208
8.1.3 The combination of place and temporal pitch cues 208

8.2 F0-discrimination by CI subjects *209*
8.2.1 Contributions of place pitch and temporal pitch to F0-discrimination 209
8.2.2 Influence of sound processing on F0-discrimination 210

8.3 An acoustic model for pitch perception by CI subjects *211*

8.4 Indications for future research *212*
8.4.1 Research platform 212
8.4.2 Pitch mechanisms in CI 213
8.4.3 Acoustic model of CI 214
8.4.4 Evaluation of the F0mod strategy 215
8.4.5 Periodicity pitch 216

8.5 Summary *218*

Bibliography **219**

List of publications **237**

Publications in international journals *237*

Conference proceedings and other publications *238*

Contributions to conferences *238*

Curriculum vitae **241**

Als doven naar muziek luisteren

Toonhoogtewaarneming met een cochleair implantaat

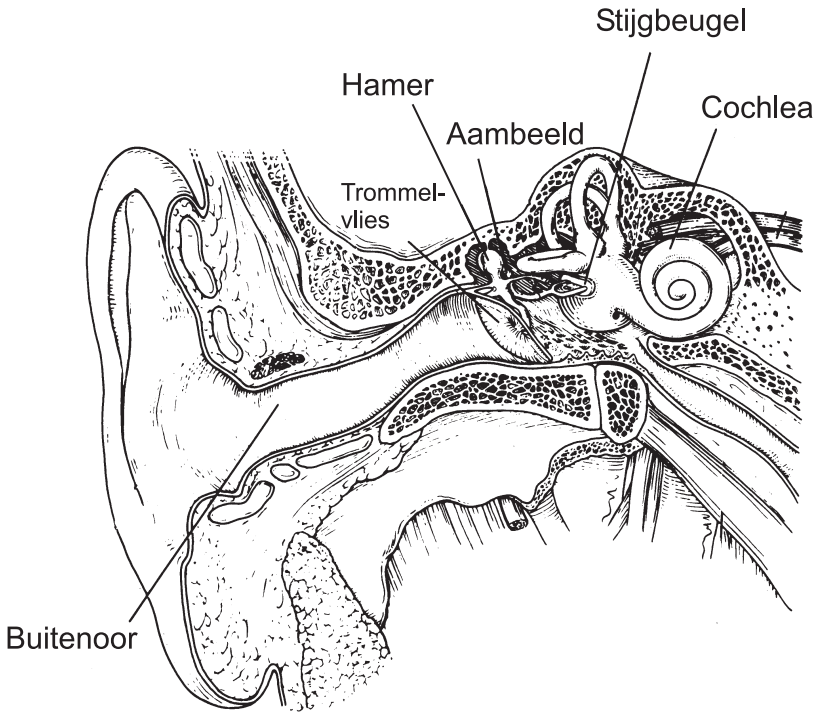
Uitgebreide Nederlandstalige samenvatting

1. Inleiding

1.1. Doof

Doof zijn is een vaak miskende en onderschatte handicap. Hoewel het doof-zijn niet zo sterk opvalt als blindheid of een fysieke handicap, zijn veel doven uitgesloten van de ‘normale maatschappij’. Doven hebben immers grote problemen om met andere mensen te communiceren. Ernstig slechthorenden kunnen zwaar lijden door dit gebrek aan menselijke interactie. Dit kan leiden tot een zwakke intellectuele ontwikkeling, een slechte job, gebrek aan eigenwaarde, onbegrip door de maatschappij (‘Hij hoort alleen wat hij wil horen!’), depressies en kwetsuren omdat ze waarschuwingssignalen niet horen. Daarom vormen de ernstig slechthorenden een soort ‘dovenmaatschappij’.

Jammer genoeg kan de medische wetenschap doofheid meestal niet genezen. Gelukkig bestaan er bijvoorbeeld gebarentaal, liplezen en hoorapparaten. Deze technieken helpen slechthorenden om met anderen te spreken, maar ze hebben hun beperkingen: gebarentaal is niet gekend door iedereen, liplezen is soms moeilijk en klassieke hoorapparaten zijn vaak onvoldoende voor zeer ernstig slechthorenden.



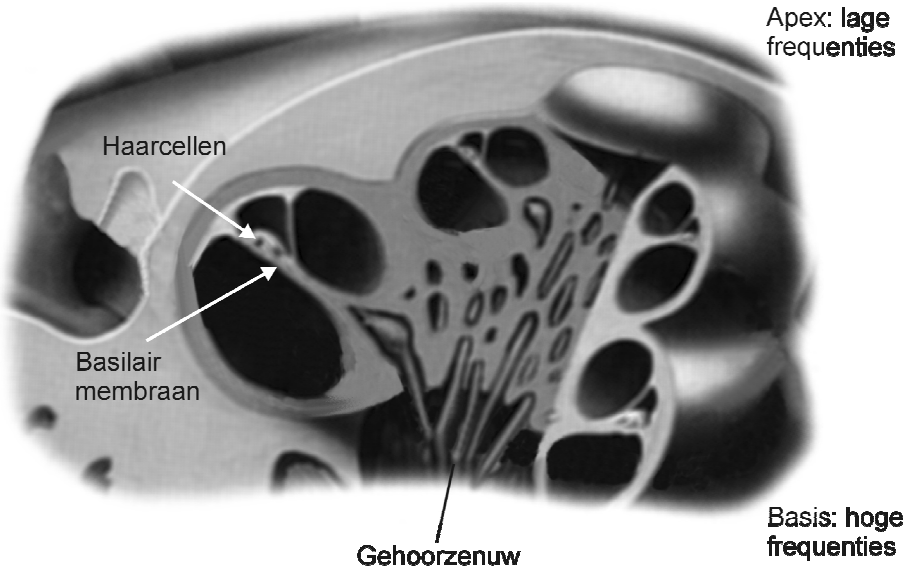
Figuur 1: De delen van het menselijke oor. Uit [Pickles, 1988].

Een klassiek hoorapparaat neemt het geluid op met een microfoontje, versterkt het geluid en stuurt het geluid uit via een mini luidsprekertje.

Een remedie voor doven of zeer ernstig slechthorenden is het cochleair implantaat (CI). Een cochleair implantaat wordt via een operatie door een chirurg in het binnenoor van de patiënt geïmplanteerd. Het CI stimuleert rechtstreeks de gehoorzenuw en slaat dus de werking van het oor over. Met behulp van een CI kan een dove voor het eerst of opnieuw horen. Met de huidige generatie van CI systemen kan de meerderheid van de geïmplanteerde patiënten spraak verstaan in een stille omgeving. De geluidservaring is echter zeer slecht als ze naar muziek luisteren. Daarom wordt in dit proefschrift getracht om de toonhoogteperceptie en de melodieperceptie bij CI gebruikers te verbeteren. We onderzoeken de mechanismen die de toonhoogte bepalen bij CI gebruikers en we ontwikkelen en testen nieuwe geluidsverwerkingsschemas. Deze nieuwe geluidsverwerkingsschemas laten de CI gebruikers toe duidelijker toonhoogte te ervaren.

1.2. Hoe werkt ons gehoor?

Geluid is de waarneming van kleine en snelle drukveranderingen door ons oor. Hoe sneller de drukveranderingen gebeuren hoe hoger het geluid klinkt. Hoe groter de

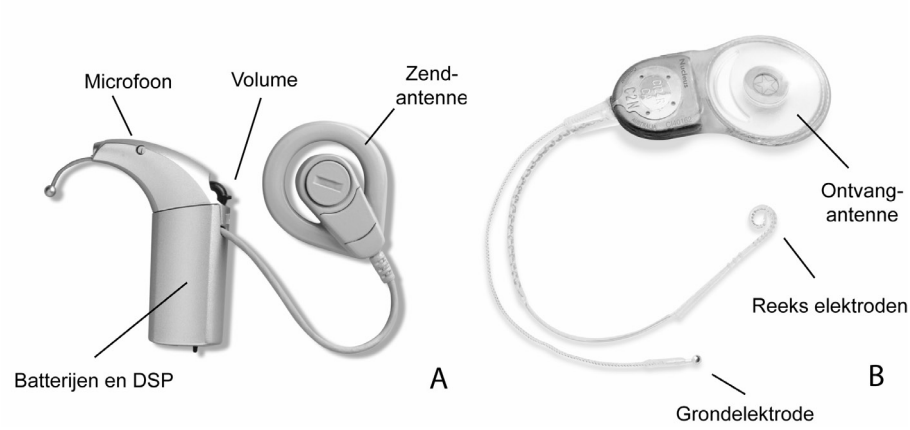


Figuur 2: Afbeelding van het slakkenhuis of cochlea. Er is een snede gemaakt om de inwendige structuur van de cochlea zichtbaar te maken.

drukveranderingen, hoe luider het geluid klinkt. Geluiden verplaatsen zich door de ruimte als drukgolven en deze drukgolven worden opgevangen door ons oor.

Figuur 1 toont de drie onderdelen van ons oor: het buitenoor, het middenoor en het binnenoor. Het buitenoor vangt de drukgolf op en geleidt deze naar het trommelvlies. Het trommelvlies begint te trillen door het geluid en de gehoorbeentjes in het middenoor geven deze mechanische beweging door aan het binnenoor. Het binnenoor omvat het slakkenhuis of cochlea.

De cochlea (zie Figuur 2) voert de belangrijkste omzetting van het geluid uit voor het naar de hersenen wordt gestuurd. De trilling van de gehoorbeentjes doet de vloeistof in de cochlea bewegen. Deze beweging wordt doorgegeven aan de vele haarcellen die zich binnen de cochlea bevinden. Het buigen van de haartjes op de haarcellen wordt omgezet in een elektrische, neurale code die doorgegeven wordt aan de hersenen door 30 000 vezeltjes van de gehoorzenuw. De snelle trillingen, met hoge frequenties en dus een hoog geluid, worden aan de basis van de cochlea doorgegeven. De trage trillingen, met lage frequenties en dus een laag geluid, worden aan de top of apex van de cochlea doorgegeven. Dit belangrijke mechanisme dat hoge van lage geluiden scheidt langs de lengte van de cochlea wordt tonotopie genoemd. De cochlea heeft de vorm van een opgerolde buis. Wanneer deze buis ‘ontrold’ zou worden heeft men de tonotopische as vanaf de top van de cochlea (apex) tot de basis van de cochlea. Op elk punt van deze as hechten vezels van de gehoorzenuw aan. Omwille van de tonotopie heeft elke vezel een karakteristieke frequentie waarbij de vezel het beste reageert. De karakteristieke frequenties in een normaalhorend persoon gaan van 20 Hz (aan de apex) tot ongeveer 20000 Hz aan de basis van de cochlea.



Figuur 3: De onderdelen van een cochleair implantaat of CI. Het linkerdeel van de figuur (A) toont het uitwendige deel dat de patiënt achter het oor draagt. Het inwendige deel dat via een chirurgische ingreep in het hoofd van de patiënt wordt gebracht, is afgebeeld in het rechterdeel (B).

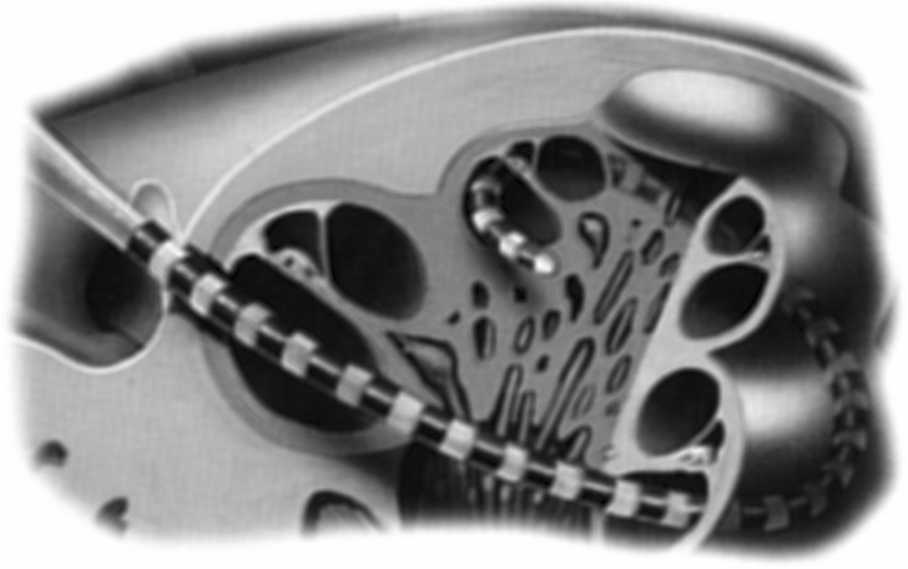
De meeste natuurlijke geluiden, o.a. spraak en muziek, bevatten zowel hoge als lage frequenties en zullen langsheen de gehele cochlea neurale responsen opwekken. De exacte verdeling van de frequenties langsheen de cochlea laat ons toe het geluid te herkennen en de toonhoogte te bepalen.

Bij een slechthorende zal minstens een van de delen van het oor niet goed werken. Vaak is het zo dat bij ernstig slechthorenden de haarcellen in de cochlea niet goed werken of zelfs afwezig zijn. Daarom is een klassiek hoorapparaat niet voldoende en is een CI noodzakelijk om het beschadigde oor te overbruggen en direct de gehoorzenuw te stimuleren.

1.3. Cochleaire implantaten

200 jaar geleden was Allesandro Volta de eerste die (eerder onaangenaam) ervaarde dat een elektrische stroom een gehoorsensatie kan teweegbrengen. Hij bracht de twee einden van een 50 Volt batterij naar ieder oor en voelde “une secousse dans la tête”, een schok in zijn hoofd, en daarna een geluid van een kokende, dikke soep. Dit ruwe principe werd verfijnd en voor de eerste keer toegepast als een remedie tegen doofheid door Djourno en Eyries [1957], die het eerste CI ontwikkelden en implanteerden.

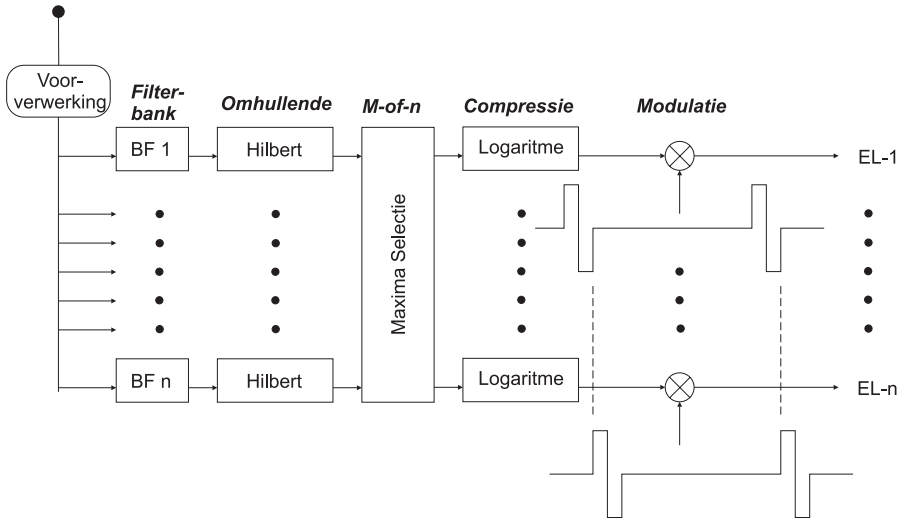
Op dit ogenblik bestaat elk CI uit twee delen: het inwendige deel, geïmplanteerd in het hoofd van de patiënt, en het uitwendige deel dat de patiënt achter het oor draagt (Figuur 3). De microfoon in het uitwendige deel pikt het geluid op. Dit geluid wordt dan gedigitaliseerd en verwerkt in de DSP (Digitale Signaal Processor) van het uitwendige deel. De DSP berekent alle elektrische stromen die naar elke elektrode van het inwendige deel moeten gestuurd worden en dit afhankelijk van de patiënt en van de binnenkomend geluid. De DSP moet ruwweg dezelfde functies doen als de delen van het oor die het CI overbrugt (buitenoor, middenoor, cochlea en haarcellen).



Figuur 4: Detail van een cochleair implantaat dat in de cochlea is gebracht.

De informatie die de DSP berekent wordt dan via een radioverbinding doorgestuurd naar het inwendige deel. Het inwendige deel bestaat uit een ontvanger en een reeks elektroden die in het slakkenhuis of cochlea geplaatst zijn. Wanneer er stroom doorgestuurd wordt zullen de zenuwvezels dicht bij die elektrode gestimuleerd worden en reageren met een neurale puls die naar de hersenen gestuurd wordt. Wanneer hoge tonen verwerkt worden zal men vooral elektroden bij de basis van de cochlea stimuleren, omdat in een normaal oor ook die regio van de cochlea zou reageren. Wanneer lage tonen verwerkt worden zullen vooral elektroden dicht bij de top van cochlea gestimuleerd worden. Op deze manier tracht men hoge en lage tonen (gedeeltelijk) te scheiden bij een CI (zie Figuur 4).

In Figuur 5 wordt de verwerking van het geluid in de DSP schematisch weergegeven. Na een voorverwerking van het geluid, wordt het geluid opgesplitst in verschillende frequentiebanden door de filterbank. Een geluid bestaat meestal uit een samenstelling van lage, midden en hoge tonen. De filterbank laat toe om die delen afzonderlijk te beschouwen en afzonderlijk aan te bieden aan de CI gebruiker op de correcte plaats in de cochlea. Elke frequentieband wordt bepaald door een banddoorlaatfilter (BF). Het aantal frequentiebanden gebruikt in dit proefschrift is 22. Dit aantal is gelijk aan het aantal elektroden dat in de cochlea geplaatst wordt bij een Nucleus CI24 implantaat. Het Nucleus CI24 implantaat wordt gebruikt voor de experimenten in dit proefschrift. Elke frequentieband komt overeen met één bepaalde elektrode. Van het signaal in elke frequentieband wordt dan de omhullende bepaald. Dit is een maat voor de intensiteit van het signaal in die frequentieband. Vervolgens gaat men alleen de frequentiebanden met de grootste intensiteit doorsturen om alleen de belangrijkste informatie over te houden (dit noemt men maxima selectie). In de compressiestap wordt het signaal aangepast aan het dynamisch bereik van de patiënt omdat CI gebruikers een zeer klein dynamisch bereik hebben vergeleken met normaalhorenden. Het signaal van elke



Figuur 5: Schematisch overzicht van de verwerking van het geluid in een CI.

frequentieband moduleert ten slotte een reeks van elektrische pulsen. De grootte van elke puls wordt dan bepaald door de intensiteit van het signaal in die bepaalde frequentieband. De elektrische pulsen worden door de elektroden in de cochlea gestuurd en deze elektrische stroom doet de gehoorzenuw reageren.

Met de recente CI systemen kunnen CI gebruikers reeds vrij goed spraak verstaan in stilte. Testen tonen aan dat gemiddeld 75% van de woorden in zinnen begrepen worden door de CI gebruikers met de huidige geluidsverwerkingsschemas. Er bestaat echter een groot verschil in resultaat tussen de patiënten. Sommige patiënten verstaan nog altijd niets met hun CI, terwijl andere zeer gemakkelijk alle spraak verstaan en zelfs de telefoon gebruiken. Voor de meeste CI gebruikers laat het CI hun toe om opnieuw deel uit te maken van de 'normale maatschappij', met een gewone job, in gewone scholen tot universitaire studies (zoals bewezen door twee proefpersonen die deelnamen aan de experimenten in dit proefschrift).

Jammer genoeg is toonhoogtewaarneming met de huidige CI systemen slecht. Zo kunnen CI gebruikers weinig melodieën herkennen ondanks het feit dat ze het ritme wel duidelijk kunnen horen. Toonhoogte is ook belangrijk voor het horen van intonatie, de herkenning van sprekers, het verstaan van spraak in tonale talen zoals Chinees, en voor het onderscheiden van meerdere simultane geluiden. In dit proefschrift wordt getracht de toonhoogtewaarneming door CI gebruikers te verbeteren.

1.4. Toonhoogte

Toonhoogte is "*dat deel van de auditieve ervaring die ons toelaat om geluiden van laag tot hoog te ordenen*". Een belangrijk deel van deze definitie is het woord 'ervaring' omdat dit aanduidt dat toonhoogte een subjectieve eigenschap is van het geluid. De toonhoogte kan dus met geen enkel instrument gemeten worden, maar is

gebonden aan de menselijke waarneming en menselijke beschrijvingen die onduidelijk en bevooroordeeld kunnen zijn.

De meeste geluiden die een toonhoogte hebben bij normaalhorende mensen, hebben een periodieke vorm van de drukgolf. Dit betekent dat de drukgolf zichzelf herhaalt elke (korte) periode. Deze geluiden noemt men harmonische geluiden. De toonhoogte die wordt waargenomen, komt meestal overeen met de frequentie waarin de drukgolf zichzelf herhaalt. Deze frequentie wordt de *fundamentele frequentie* genoemd of F0. De F0 bepaalt (bijna) eenduidig de toonhoogte. Hogere frequenties leiden tot hogere toonhoogtes en lagere frequenties tot lagere toonhoogtes. Wanneer de frequentie verdubbeld wordt, stijgt de toonhoogte met één octaaf. De frequenties van twee opeenvolgende noten (of twee naast elkaar liggende toetsen op een pianoklavier) verschillen met ongeveer 6%. Dit toonhoogteverschil wordt één *semitoon* genoemd. In een octaaf zitten exact 12 semitonen.

Toonhoogte is meer dan waarschijnlijk de belangrijkste dimensie in muziek. Toonhoogte vormt de basis voor andere muzikale begrippen zoals sleutel, akkoorden, harmonie, contour, etc. De melodie van een lied wordt voortgebracht door de variaties in toonhoogte van de noten. Wanneer de toonhoogteveranderingen wegvallen, blijft alleen het ritme over en klinkt muziek eerder saai of eentonig (zoals het woord zelf zegt). Toonhoogte is dus onontbeerlijk voor muziek in alle genres en alle culturen.

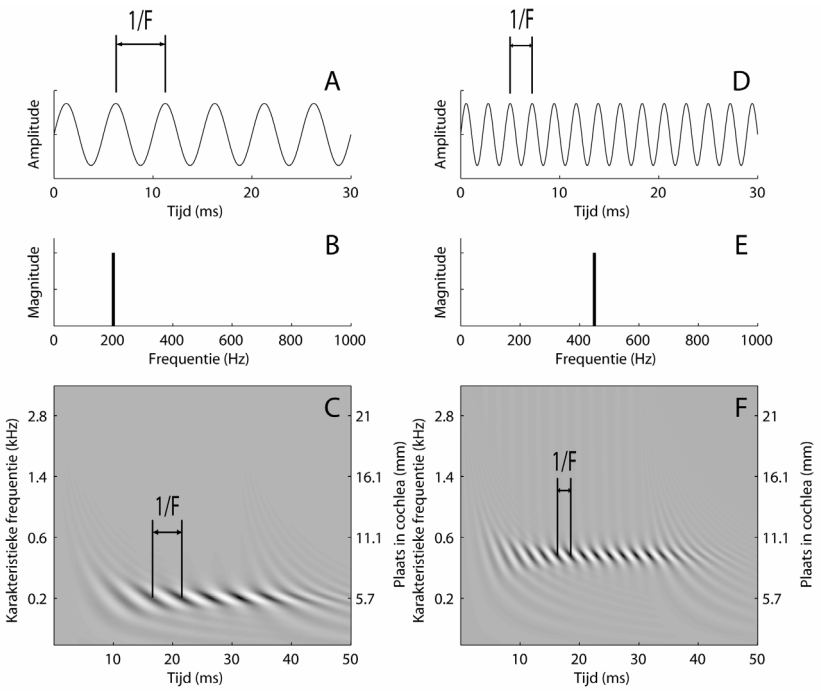
Hoewel toonhoogte niet cruciaal is voor spraakverstaan (in Westerse talen), geeft de toonhoogte extra informatie. Zo verschaft toonhoogte en toonhoogtepatronen informatie over de intonatie (of de zin een vraag of een bevestiging is), over de identiteit van de spreker (man of vrouw) en over de gevoelens van de spreker.

Toonhoogte heeft bovendien een verschillend effect voor verschillende talen. In tonale talen kan de toonhoogte de betekenis of grammaticale functie van het woord veranderen. In tonale talen kan een woord met verschillende toonhoogtepatronen worden uitgesproken, en deze verschillende toonhoogtecontouren leiden tot verschillende betekenissen. Voorbeelden van tonale talen zijn Chinees, Thais, Kantonees, Igala, … Meer dan één miljard mensen spreekt Chinees. In het Chinees bestaat bijvoorbeeld het woord ‘ma’. Naargelang de toonhoogte waarmee dit wordt uitgesproken kan dit woord onder andere moeder en paard betekenen. Voor vreemdelingen die Chinees leren kan dit dus leiden tot onverwachte en onaangename verrassingen.

Naast het feit dat toonhoogte gebruikt wordt om geluiden te ordenen van laag naar hoog, gebruiken normaalhorenden toonhoogte ook voor het onderscheiden van verschillende simultane geluiden. Zo zijn normaalhorenden in staat om een spreker te verstaan terwijl anderen in de buurt ook praten.

1.5. Neurale mechanismen voor toonhoogte

Figuur 6 geeft de golfvormen weer van twee zuivere tonen met een verschillende toonhoogte. De frequentie van de eerste toon is 200 Hz, weergeven in de linker kolom van de figuur. De tweede toon heeft een frequentie van 450 Hz en klinkt dus



Figuur 6: Afbeelding van twee zuivere tonen: de golfvorm, het spectrum en de neurale respons.

hoger, weergegeven in de rechterkolom van de figuur. Het spectrum van de twee tonen is weergegeven in de middelste rij.

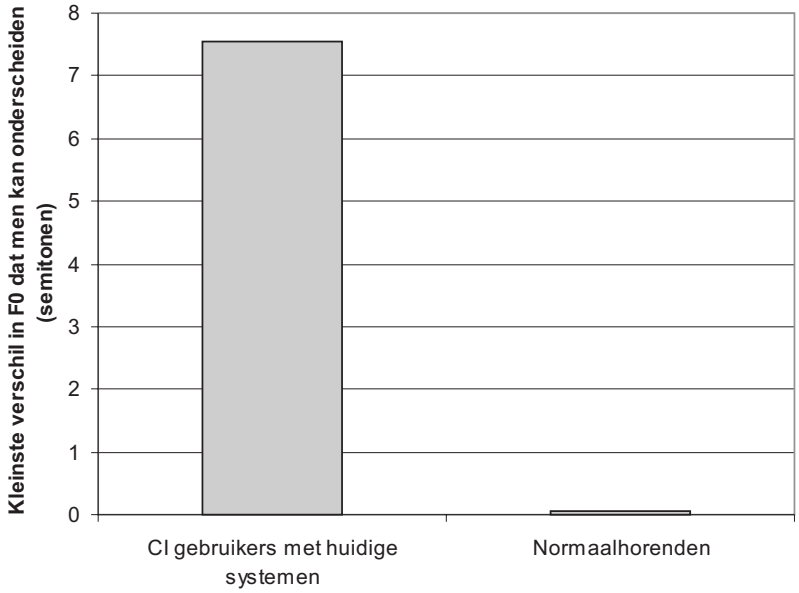
In de onderste rij van Figuur 6 is de neurale respons weergegeven voor de twee tonen. De neurale respons is het grootste in de buurt van de zenuwvezels met karakteristieke frequentie gelijk aan die van de aangeboden toon. Hoe verder men zich van deze plaats in de cochlea verwijdert (langs de y-as in de figuur), hoe kleiner de respons. De neurale respons herhaalt zich ook in de tijd met dezelfde frequentie als de frequentie van de aangeboden toon. Dit betekent dat de respons groter is voor elke 'top' van de zuivere toon en dat de respons kleiner is voor elke 'dip' van de zuivere toon.

Wanneer de neurale responsen voor beide zuivere tonen vergeleken worden zijn er twee duidelijke verschillen:

1. Bij een hogere toon is de neurale respons meer gelegen naar de basis toe. Dit is dus een plaatsafhankelijke component.

2. Bij een hogere toon varieert de neurale respons sneller in de tijd, de pieken en dalen in de neurale respons volgen elkaar sneller op. Dit is een tijdsafhankelijke component.

Deze twee componenten vormen de basis van de moderne theorieën over het mechanisme voor toonhoogte. Theorieën over toonhoogte bestaan als sinds de snaar



Figuur 7: Het kleinste verschil in toonhoogte dat normaalhorden en CI gebruikers kunnen onderscheiden. Kleiner is beter.

theorie van Pythagoras (600 v. Chr.). Nieuwere theorieën werden ontwikkeld doorheen de daaropvolgende eeuwen, maar er bestaat nog steeds een heftig (en soms vurig) debat tussen de onderzoekers over hoe toonhoogte wordt waargenomen. Momenteel wordt echter door de meeste onderzoekers aangenomen dat de toonhoogteperceptie gebaseerd is op kleine veranderingen in zowel de plaats-component als in de tijdscomponent.

1.6. Toonhoogte en muziek bij CI gebruikers

De meeste onderzoeken met CI gebruikers concentreerden zich op spraak en spraakverstaan omdat het CI vooral tot doel heeft de gebruikers te laten communiceren. Het is slechts de laatste jaren dat toonhoogte en muziek bij CI gebruikers onderzocht werd in grotere groepen van proefpersonen. Uit deze studies blijkt dat de meeste CI gebruikers muziek slechts weinig appreciëren. Dit komt vooral omdat de melodie slecht of zelfs niet wordt waargenomen door de CI gebruikers. Het ritme van de muziek kunnen de CI gebruikers bijna zo goed als normaalhorenden horen. Daarom luisteren de meeste CI gebruikers het liefst naar muziek met een zeer uitgesproken ritme en hoeft de muziek voor hen niet noodzakelijk veel melodie te bevatten.

Dit probleem van CI gebruikers om de contouren in een melodie te volgen wordt zeer waarschijnlijk veroorzaakt door de gebrekkige perceptie van toonhoogte door CI gebruikers. Het kleinste verschil in toonhoogte dat CI gebruikers kunnen

waarnemen verschilt zeer sterk van patiënt tot patiënt, en varieert van 1 semitoon tot 24 semitonen met een gemiddelde van ongeveer 7 semitonen. Dit betekent dat twee noten die minder dan 7 toetsen uit elkaar liggen op een pianoklavier niet kunnen worden onderscheiden door CI gebruikers. Dit onderscheidingsvermogen is veel slechter dan bij normaalhorenden. Normaalhorenden kunnen ongeveer nog 10 stappen in toonhoogte onderscheiden tussen twee naast elkaar liggende toetsen op het pianoklavier. Het kleinste verschil dat normaalhorenden kunnen onderscheiden in toonhoogte bedraagt ongeveer 0.1 semitoon of ongeveer 50 keer beter dan CI gebruikers (Figuur 7).

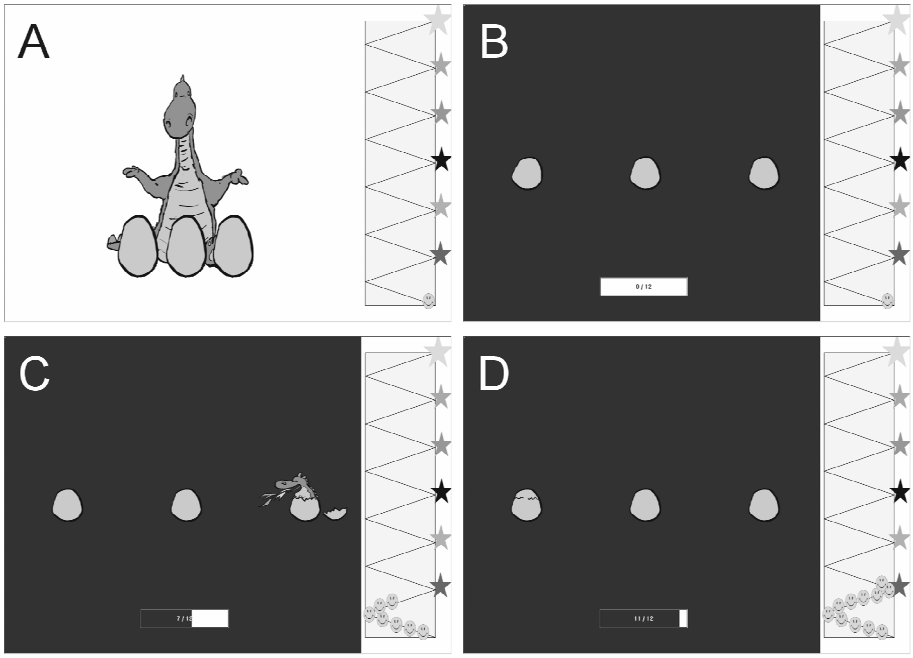
Bij CI gebruikers bestaan twee mechanismen waarmee toonhoogteverschillen waargenomen worden: temporele toonhoogte en plaatstoonhoogte. Wanneer men de stroom door een bepaalde elektrode steeds sneller doet variëren, hoort de CI gebruiker een opgaande toon. Dit noemt men *temporele toonhoogte*. Wanneer men achtereenvolgens de stroom stuurt naar elektroden die variëren in locatie van de apex van de cochlea naar de basis, dan hoort de CI gebruiker ook een opgaande toon. Dit noemt men *plaatstoonhoogte*. De twee mechanismen voor toonhoogte in CI gebruikers zijn sterk gerelateerd aan de twee componenten in de neurale respons van normaalhorenden voor toonhoogteverschillen (Zie Figuur 6).

1.7. Overzicht van dit proefschrift

In de voorafgaande paragrafen heb ik de stand van zaken voor cochleaire implantaten beknopt uiteengezet. Ook gaf ik een overzicht van de *state-of-the-art* over toonhoogteperceptie. Deze inleiding is vooral gebaseerd op informatie uit studies die aan dit proefschrift vooraf gingen. Deze informatie vormt de wetenschappelijke basis voor de eigen en nieuwe bijdragen uit dit proefschrift. De voorafgaande paragrafen worden gedetailleerder uitgewerkt in hoofdstuk 1 van dit proefschrift.

In het vervolg van deze Nederlandstalige samenvatting beschrijf ik de belangrijkste bevinden die ik bekomen heb in dit proefschrift. Hierbij komt elke genummerde sectie overeen met het respectievelijke hoofdstuk uit het proefschrift (bvb. Sectie 2 is de samenvatting van Hoofdstuk 2). De belangrijkste doelstellingen van dit proefschrift waren tweeledig. Ten eerste, het beter begrijpen van de toonhoogtemechanismen bij CI gebruikers; en ten tweede het gebruiken van deze kennis om de toonhoogtewaarneming bij CI gebruikers te verbeteren door middel van nieuwe geluidsverwerkingsschemas. De belangrijkste eigen en nieuwe ontwikkelingen die tot stand gekomen zijn vanuit deze doelstellingen zijn:

- Een nieuw onderzoeksplatform dat toelaat om Nucleus CI24 gebruikers te testen (Hoofdstuk 2).
- Een beter begrip van het mechanisme voor plaatstoonhoogte (Hoofdstukken 3 en 5) en van het mechanisme voor temporele toonhoogte (Hoofdstukken 5 en 7).
- Een beter begrip van hoe plaatstoonhoogte en temporele toonhoogte gecombineerd worden in onze hersenen (Hoofdstuk 4).



Figuur 8: Overzicht van verschillende schermen voor de kinderversie van het onderzoeksplatform. Paneel A geeft een afbeelding weer uit een inleidend filmpje dat de taak uitlegt aan de kinderen. Panelen B tot D geven de interface weer tijdens het experiment. Elk eitje op het scherm komt overeen met een bepaald geluid. Als de kinderen het juiste eitje aanduiden springt er een draakje uit het ei (Paneel C). Als het kind het verkeerde eitje aanduidt breekt het eitje maar verschijnt er geen draakje (Paneel D).

- De ontwikkeling en evaluatie van een simulatie van het geluid dat CI gebruikers horen. Dankzij deze simulatie kunnen normaalhorende personen een idee krijgen van wat CI gebruikers horen (Hoofdstuk 6).
- De evaluatie van het effect van de huidige geluidsverwerkingsschemas op de toonhoogtewaarneming bij CI gebruikers (Hoofdstuk 5).
- De ontwikkeling en evaluatie van een nieuw geluidsverwerkingsschema dat de toonhoogtewaarneming en de muziekervaring van CI gebruikers significant verbetert (Hoofdstuk 7).

2. Het onderzoeksplatform: APEX

Omdat toonhoogte per definitie een subjectieve ervaring van de mens is, kan toonhoogte alleen onderzocht worden door middel van experimenten met mensen. Meer specifiek is het in dit proefschrift noodzakelijk om experimenten te doen met gebruikers van een cochleair implantaat. Daarom ontwikkelen we in het begin van dit proefschrift een onderzoeksplatform dat toelaat om gebruikers van een CI makkelijk te testen en hun responsen te analyseren.

Het ontwikkelde onderzoeksplatform, APEX genoemd, laat de onderzoeker toe om op een gebruiksvriendelijke wijze willekeurige experimenten op te zetten. Het onderzoeksplatform zorgt voor de automatische aanbieding van de stimuli en voor het verzamelen en analyseren van de antwoorden van de proefpersonen. De opstelling is in staat om elektrisch te stimuleren via een CI en om akoestisch te stimuleren. Op dit ogenblik is APEX in staat om drie verschillende CI types aan te sturen: het LAURA implantaat, het Nucleus CI22 implantaat en het Nucleus CI24 implantaat. Alle CI gebruikers die als proefpersoon meewerken aan de experimenten in dit proefschrift zijn gebruikers van het Nucleus CI24 implantaat. Dit implantaat heeft de meeste elektroden van alle huidige commerciële CI systemen.

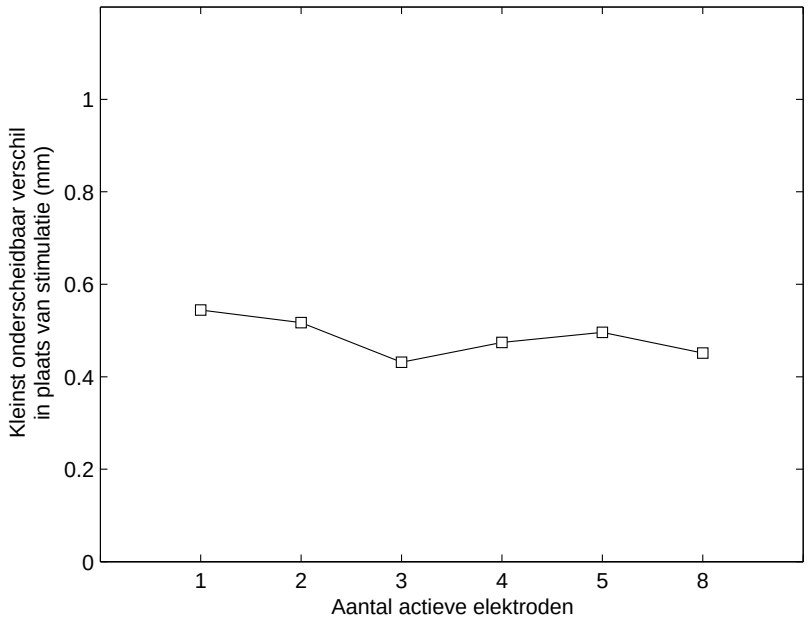
De grafische vormgeving van het programma is ook aangepast zodat het mogelijk is om jonge kinderen te testen. Met behulp van animaties en tekeningen kunnen reproduceerbare en betrouwbare metingen bekomen worden van kinderen vanaf ongeveer 2.5 jaar oud. De experimenten worden hierdoor als het ware in een soort computerspel ingebed (zie Figuur 8). De kinderinterface is ontwikkeld via een samenwerking met het departement pedagogie.

Het onderzoeksplatform is gebruikt in alle experimenten die deel uitmaken van dit proefschrift. Daarnaast werd het onderzoeksplatform ook in tal van andere studies binnen de KULeuven gebruikt alsook binnen andere universiteiten over de hele wereld (Nederland, Verenigd Koninkrijk, Nieuw Zeeland, Verenigde Staten, Zwitserland, …). APEX is gratis beschikbaar voor andere onderzoeksgroepen.

3. Plaatstoonhoogte met stimulatie van meerdere elektroden

CI gebruikers merken een opgaande toonhoogte als de plaats van de elektrische stimulatie in de cochlea verplaatst wordt van de apex naar de basis. Dit is de plaatstoonhoogte, zoals eerder vermeld. De meeste studies tot nu toe hebben plaatstoonhoogte onderzocht met stimuli die slechts één elektrode activeren. Nochtans, alle huidige CI systemen activeren meerdere elektroden gelijktijdig of zeer snel afwisselend. Daarom onderzoeken we in dit experiment hoe de gevoeligheid voor plaatstoonhoogte varieert als meer dan één elektrode wordt geactiveerd.

We meten het kleinste verschil op dat de CI gebruikers nog net kunnen onderscheiden naargelang de plaats van stimulatie. We doen dit bij 4 CI gebruikers en in functie van het aantal geactiveerde elektroden (van 1 tot en met 8). We bieden telkens twee verschillende geluiden aan en vragen de proefpersonen aan te duiden welk geluid het hoogste was. Wanneer het verschil in toonhoogte groot is zullen de proefpersonen het telkens juist hebben. Wanneer het verschil echter zeer klein wordt zullen de proefpersonen af en toe fouten beginnen maken. Het kleinste verschil dat de CI gebruikers nog kunnen onderscheiden (Kleinst onderscheidbaar verschil KOV) in toonhoogte wordt gedefinieerd door het verschil in toonhoogte waarbij de proefpersonen nog 75% van de keren correct het hoogste geluid van de twee aanduiden.



Figuur 9: De gevoeligheid voor verschillen in plaatstoonhoogte in functie van het aantal geactiveerde elektroden.

De resultaten van het experiment tonen aan dat het aantal geactiveerde elektroden geen significant effect heeft op de gevoeligheid van de CI gebruikers voor plaatstoonhoogte (Zie Figuur 9). Dit resultaat komt overeen met twee modellen voor plaatstoonhoogte. Volgens het eerste model worden de flanken van de stimulatiepatronen vergeleken tussen de twee geluiden door de CI gebruikers. Volgens het tweede model vergelijken de CI gebruikers (onbewust) de zwaartepunten van de stimulatiepatronen van de twee geluiden. De gevoeligheid van de proefpersonen verschilt van proefpersoon tot proefpersoon. Het kleinste verschil dat ze kunnen onderscheiden op basis van plaatstoonhoogte varieert van 0.25 tot 0.46 mm.

4. De combinatie van plaatstoonhoogte en temporele toonhoogte

Zoals reeds gezegd bestaan er bij CI gebruikers twee mechanismen voor toonhoogte: temporele toonhoogte en plaatstoonhoogte. Deze twee mechanismen wekken beide een bepaalde toonhoogte op, maar toch zijn de sensaties komend uit die twee mechanismen niet identiek. Beide toonhoogtesensaties kunnen onafhankelijk van elkaar gevarieerd worden. De meeste studies hebben plaats- en temporele toonhoogte gescheiden van elkaar onderzocht. Slechts weinig studies hebben getoetst hoe de toonhoogte ervaren wordt wanneer beide toonhoogtemechanismen simultaan werken. In dit experiment wordt onderzocht hoe proefpersonen plaats en temporele toonhoogte combineren in een ‘globale’ toonhoogtesensatie.

In het experiment bieden we normaalhorenden een simulatie aan van wat CI gebruikers horen. Het ontwerp van deze simulatie is uitgewerkt in hoofdstuk 6 van dit proefschrift. We bieden de proefpersonen twee geluiden aan die zowel in plaatstoonhoogte als in temporele toonhoogte van elkaar verschillen. Het verschil in plaatstoonhoogte en temporele toonhoogte kan consistent of inconsistent zijn. Consistent betekent dat beide toonhoogteverschillen een stijging van de toonhoogte opwekken, of beide wekken een daling van de toonhoogte op. Inconsistent betekent dat één mechanisme een dalend toonhoogteverschil opwekt, terwijl het andere mechanisme een stijgend toonhoogteverschil opwekt.

De resultaten van het experiment tonen aan dat de twee mechanismen voor toonhoogte geïntegreerd worden in een gecombineerde toonhoogtegewaarwording. Deze combinatie is significant verschillend voor consistente en inconsistente toonhoogteverschillen. Bij consistente verschillen in toonhoogte versterkt het toonhoogteverschil van het ene mechanisme het toonhoogteverschil van het andere mechanisme, en is er dus constructieve integratie van beide mechanismen. Bij inconsistente toonhoogteverschillen kan het ene toonhoogteverschil het toonhoogteverschil van het andere mechanisme te niet doen, en is er dus sprake van destructieve interferentie.

Uit de resultaten leiden we een formule af die de duidelijkheid van het gecombineerde toonhoogteverschil voorspelt op basis van het zuivere temporele toonhoogteverschil en het zuivere plaatsafhankelijke toonhoogteverschil.

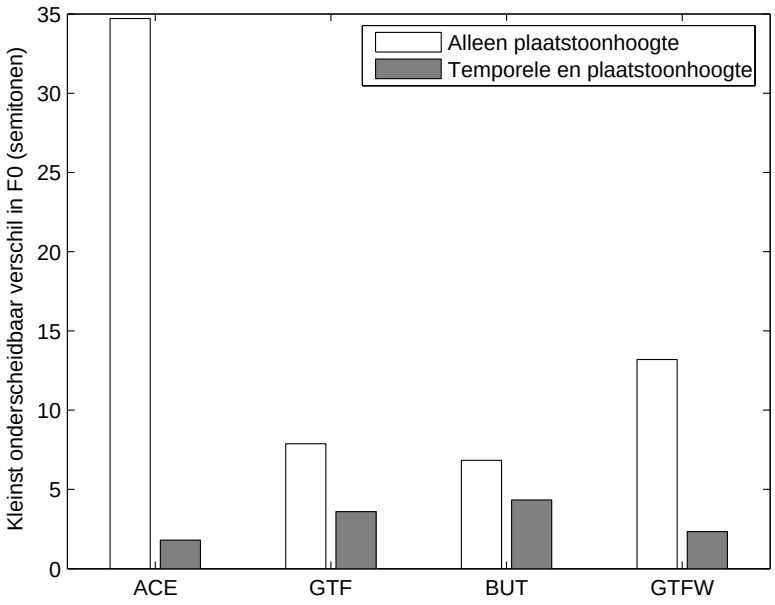
$$\text{sign}(d'_{gecombineerd})d'^{2}_{gecombineerd} = \text{sign}(d'_{plaats})d'^{2}_{plaats} + \text{sign}(d'_{temporeel})d'^{2}_{temporeel}$$

In deze formule zijn de bijdragen van beide mechanismen voor toonhoogte onafhankelijk van elkaar en wordt elke bijdrage optimaal geïntegreerd in de gecombineerde toonhoogtegewaarwording. Deze formule wordt in het vervolg van dit proefschrift gebruikt.

5. Het effect van de filterbank op het onderscheiden van F0

De DSP in het CI systeem transformeert het binnenkomende geluid in een complex elektrisch stimulatiepatroon dat alle elektroden activeert. Daartegenover staat dat de meeste onderzoeken enkel synthetische signalen gebruikt die niet voortkomen uit een akoestisch geluid uit het dagelijkse leven. De meeste studies, inclusief de vorige experimenten uit dit proefschrift, gebruiken synthetische artificiële stimuli die geen directe relatie hebben tot de geluiden om ons heen. In dit experiment onderzoeken we de toonhoogtewaarneming van CI gebruikers voor gestileerde klinkers die door de DSP van het CI systeem verwerkt zijn. Meer specifiek onderzoeken we het effect van de filterbank in DSP op het kunnen onderscheiden van kleine toonhoogteverschillen.

We meten het kleinst onderscheidbaar verschil in toonhoogte tussen twee klinkers op en dit voor vier verschillende filterbanken: ACE, GTF, BUT en GTFW. De banddoorlaatfilters van deze vier filterbanken verschillen in de positie van de frequentiebanden, de overlap tussen de frequentiebanden en de vorm van de



Figuur 10: Kleinste verschil in toonhoogte dat de CI gebruikers kunnen onderscheiden. Kleiner is beter.

frequentiebanden. Bovendien wordt alles ook onder twee condities getest om de bijdragen van plaatstoonhoogte en temporele toonhoogte apart te beoordelen. In de eerste conditie is alleen het mechanisme voor plaatstoonhoogte actief omdat de omhullende van elk kanaal stationair gemaakt is. In de tweede conditie zijn zowel het mechanisme voor temporele toonhoogte als het mechanisme voor plaatstoonhoogte actief.

De resultaten zijn weergegeven in Figuur 10 en tonen aan dat het onderscheiden van F0 op basis van alleen plaatstoonhoogte mogelijk is maar dat het onderscheidend vermogen zeer zwak is. Het kleinst onderscheidbare verschil kan oplopen tot meer dan 1 octaaf, afhankelijk van de gebruikte filterbank. Het activeren van het temporele toonhoogte mechanisme verbetert zeer sterk het vermogen van de CI gebruikers om toonhoogte verschillen te detecteren. De kleinste verschillen in toonhoogte die de CI gebruikers kunnen waarnemen liggen tussen 2 en 4 semitonen.

Een analyse van de resultaten toont aan dat de resultaten die alleen gebaseerd zijn op plaatstoonhoogte overeenkomen met een model dat het zwaartepunt van de elektrische stimulatiepatronen van geluiden vergelijkt. Het geluid met het zwaartepunt dat het meest naar de basis van de cochlea ligt, klinkt het hoogst. De analyse van de resultaten waarbij het mechanisme voor temporele toonhoogte geactiveerd is, toont aan dat filterbanken met relatief brede frequentiebanden de beste toonhoogte-informatie verschaffen aan de CI gebruikers. Het mechanisme voor temporele toonhoogte werkt het best wanneer de modulaties in de tijd in

verschillende kanalen zo groot mogelijk zijn en wanneer deze modulaties ook synchroon gebeuren over de verschillende kanalen.

Jammer genoeg moeten we besluiten uit dit experiment dat er geen enkele filterbank is die zowel het mechanisme voor plaatstoonhoogte als het mechanisme voor temporele toonhoogte optimaal benut. Daarom is het nodig om nieuwe geluidsverwerkingsschemas te ontwerpen die specifiek toonhoogte trachten te verbeteren. In hoofdstuk 7 van dit proefschrift wordt dan ook een nieuw geluidsverwerkingsschema ontwikkeld en geëvalueerd dat de toonhoogte-waarneming en de muziekervaring van CI gebruikers sterk verbetert.

6. Simulatie van de toonhoogte van een CI geluid

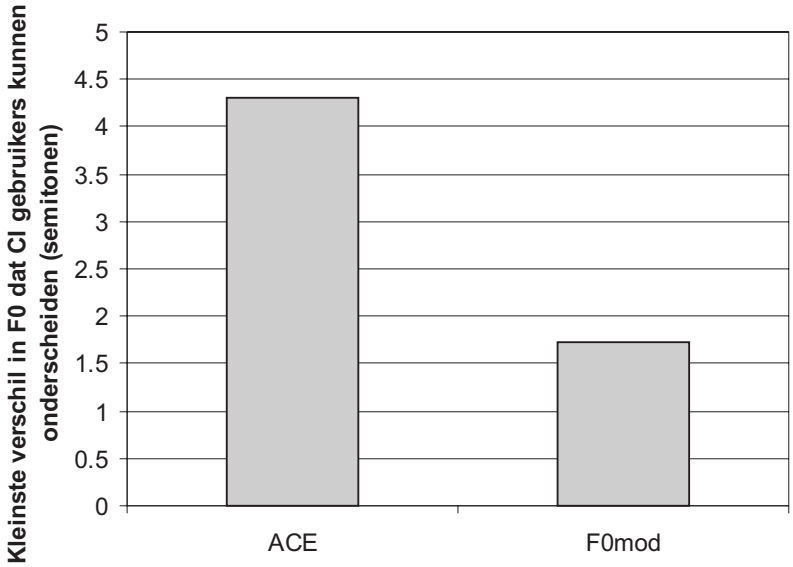
In hoofdstuk 6 van dit proefschrift ontwikkelen en testen we een simulatie van het geluid dat een CI gebruiker hoort. Zulke een simulatie kunnen we aanbieden aan normaalhorenden en zo kunnen die ervaren wat de geluidskwaliteit is voor een CI gebruiker. Hierbij moet echter opgemerkt worden dat deze simulatie niet exact hetzelfde is geluid als het geluid dat een CI gebruiker hoort. Er zijn verschillen tussen het geluid dat een CI gebruiker hoort en de simulatie die ermee samengaat. Men kan de simulatie beter beschouwen als een voorstelling van de beperkte geluidsinformatie die CI gebruikers verkrijgen, en geen volledige afspiegeling van het CI geluid.

Het gebruik van een simulatie heeft verschillende voordelen. Alle normaalhorenden hebben bijvoorbeeld vergelijkbare resultaten bij eenzelfde test, terwijl dit niet altijd zo is bij CI gebruikers omdat hun auditief systeem sterk kan verschillen omwille van verschillende oorzaken van doofheid. Ook kan men in een simulatie parameters veranderen die onmogelijk veranderd kunnen worden bij CI gebruikers (zoals het aantal elektroden of de plaats van de elektroden) en kunnen op deze manier nieuwe CI systemen getest worden voor ze geïmplanteerd worden in een patiënt. Simulaties van CI geluiden geven de onderzoeker, of de ouder van een geïmplanteerd kind, ook de kans om te horen hoe (relatief slecht) het geluid klinkt voor een CI gebruiker.

Simulaties van CI geluiden werden reeds gebruikt in andere studies om het spraakverstaan van CI gebruikers te onderzoeken. Deze studies over het verstaan van spraak toonden aan dat resultaten bekomen met deze simulaties (de standaard CI simulatie genoemd) goed overeenkomen met resultaten die opgemeten worden bij CI gebruikers. In deze studie gaan we na of we de standaard simulatie ook kunnen gebruiken voor het onderzoek van toonhoogte. Meer specifiek gaan we na of de resultaten die we bekomen hebben in de vorige hoofdstukken van dit proefschrift nagebootst kunnen worden met een CI simulatie bij normaalhorenden.

De evaluatie gebeurt voor twee CI simulaties: de standaard CI simulatie (zoals gebruikt in vorige studies) en een nieuwe CI simulatie (de CISIM simulatie). In deze simulatie wordt rekening gehouden met de geluidsverwerking uit de DSP van het CI systeem en met de fysiologie van elektrische stimulatie in de cochlea.

De resultaten tonen aan dat resultaten bekomen met de standaard CI simulaties (zoals gebruikt in vorige studies) niet overeenkomen met resultaten die opgemeten



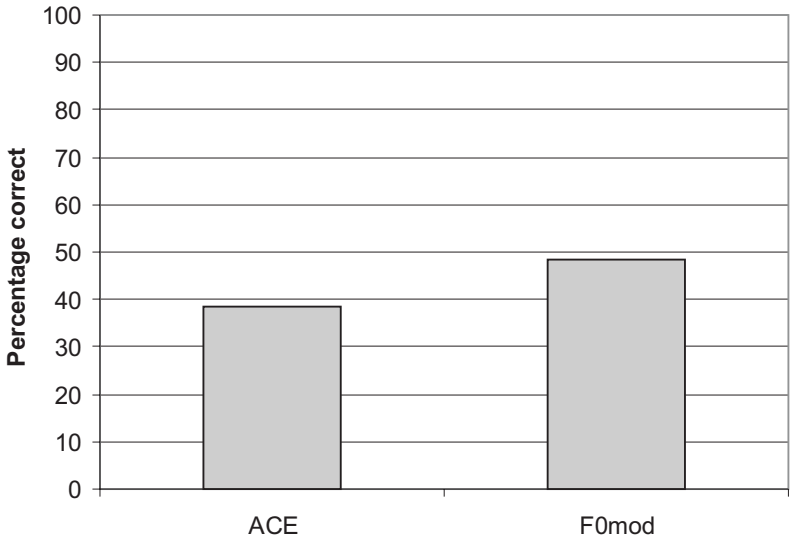
Figuur 11: Kleinste verschil in toonhoogte dat de CI gebruikers kunnen onderscheiden voor noten onder 300 Hz van 5 verschillende muziekinstrumenten. Kleiner is beter.

werden in CI gebruikers. Daarentegen komen de resultaten van onze nieuwe CISIM simulatie van CI geluiden wel overeen met resultaten van CI gebruikers.

7. Verbeterde muziek door nieuwe F0mod geluidsverwerking

CI gebruikers hebben zeer grote problemen met de waarneming van toonhoogte. CI gebruikers horen bijna geen verschil tussen twee tonen die normaalhorenden gemakkelijk kunnen onderscheiden. Dit probleem wordt onder andere veroorzaakt door de geluidsverwerking in de DSP die toonhoogte niet accuraat doorstuurt. In hoofdstuk 5 van dit proefschrift tonen we aan dat met de huidige generatie van geluidsverwerkingsschemas het onmogelijk is om optimaal toonhoogte informatie door te sturen, zelfs niet als de filterbank of de omhullende-detectie veranderd wordt.

In hoofdstuk 7 van dit proefschrift stellen we een nieuw geluidsverwerkingsschema voor dat ontworpen is om specifiek de waarneming van toonhoogte en muziek door CI gebruikers te verbeteren. Dit nieuwe schema wordt het F0mod schema genoemd en is gebaseerd op de resultaten van de vorige hoofdstukken. Het nieuwe schema tracht zowel het mechanisme voor plaatstoonhoogte als het mechanisme voor temporele toonhoogte optimaal te benutten. Daarom wordt in het F0mod schema de uitgang van de filterbank expliciet gemoduleerd aan de F0 frequentie. Deze modulatie gebeurt synchroon over de verschillende kanalen en met de maximale



Figuur 12: Melodieherkenning van eenvoudige Vlaamse liedjes.

modulatiediepte. Dit optimaliseert het stimulatiepatroon voor het temporele toonhoogtemechanisme. De filterbank die gebruikt wordt is geoptimaliseerd voor het mechanisme voor plaatstoonhoogte.

Het nieuwe F0mod schema is geëvalueerd in vier verschillende experimenten. In elk experiment worden de resultaten van het nieuwe schema vergeleken met resultaten van het momenteel meest gebruikte schema (ACE). De resultaten van de eerste twee experimenten tonen aan dat het onderscheiden van toonhoogtes voor klinkers niet significant verschillend is voor het nieuwe schema, maar dat drie maal kleinere verschillen in toonhoogte kunnen onderscheiden worden met het nieuwe F0mod schema voor muziekinstrumenten (zie Figuur 11).

In het derde experiment bieden we de CI gebruikers eenvoudige Vlaamse melodietjes aan en vragen hen de titel van het melodietje aan te duiden. De CI gebruikers herkennen meer melodietjes met het nieuwe F0mod schema dan met het huidig gebruikte geluidsverwerkingsschema (zie Figuur 12). In het vierde experiment wordt een muzikaal opgeleide CI gebruiker gevraagd om de toonhoogte-intervallen, of afstanden tussen twee noten in toonhoogte, te benoemen. Uit de resultaten van dit experiment blijkt dat met het nieuwe F0mod schema deze CI gebruiker minder grote fouten maakt dan met het huidige geluidsverwerkingsschema.

We besluiten dus in hoofdstuk 7 dat het nieuw ontworpen F0mod schema zowel de waarneming van toonhoogte als de waarneming van muziek verbetert.

8. Besluiten

Vergeleken met normaalhorenden ervaren CI gebruikers toonhoogte en verschillen in toonhoogte zeer zwak. Normaalhorenden kunnen zeer kleine verschillen in toonhoogte onderscheiden van 0.1 tot 0.2 semitonen. CI gebruikers kunnen daarentegen slechts verschillen in toonhoogte van 7 semitonen onderscheiden. Dit vermogen om toonhoogtes te onderscheiden is te zwak om muziek te appreciëren. Daarom werd in dit proefschrift getracht de gewaarwording van toonhoogte bij CI gebruikers te verbeteren. In een eerste deel experimenten onderzoeken we de mechanismen die gebruikt worden voor toonhoogte bij CI gebruikers en in het tweede deel van de experimenten gebruiken we deze informatie om de verwerking van het geluid in een CI systeem te verbeteren zodat muziek en toonhoogte beter waargenomen wordt.

We bouwen een onderzoeksplatform dat ons toelaat om verscheiden experimenten uit te voeren met CI gebruikers. De resultaten van deze experimenten tonen aan dat het mechanisme voor plaatstoonhoogte bij CI gebruikers gerelateerd is aan het zwaartepunt van het stimulatiepatroon. Dit betekent dat plaatstoonhoogte overeenkomt met spectrale toonhoogte bij normaalhorenden of met de scherpte van de klankkleur. De resultaten tonen ook aan dat de gevoeligheid van plaatstoonhoogte bepaald wordt door de steilheid van de flanken van het stimulatiepatroon. We demonstreren ook dat het mechanisme voor temporele toonhoogte het best werkt wanneer de modulaties over de kanalen synchroon zijn en de modulatiediepte maximaal is.

Gebaseerd op deze informatie hebben we een nieuw schema ontwikkeld voor de verwerking van het geluid in een CI systeem dat de toonhoogtewaarneming verbetert: het F0mod schema. Dankzij dit nieuwe schema kunnen de CI gebruikers drie keer kleinere toonhoogteverschillen onderscheiden en herkennen ze meer melodieën.

Chapter 1

Introduction to cochlear implants and pitch perception

This chapter introduces cochlear implants and pitch perception. It describes what cochlear implants are, and how they are an effective means to, at least partially, overcome deafness. This chapter also describes the basic concepts associated with pitch perception and the state-of-the-art on how pitch is perceived in normal-hearing subjects, and secondly in subjects fitted with a cochlear implant. Pitch is important for intonation, melody, and the meaning of words in tonal languages such as Mandarin Chinese. The chapter then highlights that although relatively good results have been obtained with cochlear implants for speech understanding, all patients implanted with a cochlear implant exhibit poor pitch perception. The improvement of this poor pitch perception with current cochlear implant devices is the main topic of this thesis. The chapter concludes with an overview of the major research questions of this thesis.

1.1 Cochlear implants

1.1.1 Hearing loss

Hearing loss is widely recognized as one of the most common human disorders. On average one out of thousand newborns is affected by a severe hearing loss either congenital or acquired. Moreover, the prevalence of hearing loss increases monotonically for older populations as the patient's hearing is irreversibly affected

by, for instance, noise induced trauma and age-related hair-cell degeneration. As a result, about half of the people of 65 and older suffer from a mild to severe hearing loss.

Deafness is an often underestimated and misjudged handicap that seriously limits patients' capabilities in life. Although the handicap of the deaf is not as conspicuous as for instance blindness or a physical handicap, deaf people find themselves often excluded from the 'normal' society because of the large problems they have in communicating with other people. Severely hearing impaired people suffer greatly from this lack of human interaction. This can possibly lead to poor intellectual development, problems in their professional lives, low self-esteem, stigmatization by the community ("He only hears what he wants to hear!"), depression, injuries from accidents from failing to hear warning signals, … As a result, severely hearing impaired people often seek comfort with each other and help each other out, forming a 'deaf community'.

Unfortunately, medical science is not capable to cure deafness, at least for most of the affected patients. On the other hand a number of techniques exist that enable the patients to communicate, such as lip-reading, sign language and of course hearing aids. Most of these techniques are of limited help because for instance lip-reading only provides a small part of the information the spoken word gives and sign language is only known by a relatively small amount of people. Classical hearing aids, which pick up and amplify the sound before returning it to the ear through a kind of miniature loudspeaker, are an effective rehabilitative means for mild to moderate hearing losses. However, for subjects affected by profound hearing loss classical hearing aids provide insufficient speech understanding and for these subjects a cochlear implant may provide a means to communicate with the world. Cochlear implants are a special kind of hearing device that is surgically implanted inside the head of the patient. It bypasses the ear and provides sound sensations to the patient through direct electrical stimulation of the auditory nerve.

1.1.2 How does hearing work?

Sound perception is the sensation of fluctuations in the air pressure. The amplitude of this pressure wave determines the loudness and its frequency determines the perceived pitch. A young normal person can hear sounds with frequencies between 20 Hz and 20000 Hz. Sounds are, however, never composed of one single frequency but contain multiple frequencies at the same time.

Sounds travel through the air as pressure waves and are captured by the ear which transforms this pressure wave into a neural code which is interpreted by our brain. Our ear basically consists of three parts: the outer ear, the middle ear and the inner ear or cochlea (see Figure 1.1). The outer ear consists of the auricle (pinna and concha), the ear canal (auditory meatus) and the ear drum (tympanic membrane). The main function of the outer ear is to capture the pressure wave and to let it travel towards the ear drum. The outer ear works like a horn and amplifies the pressure wave approximately 15 dB around 3 kHz and also changes the timbre of the sound depending on the position of the sound source. The ear drum is taut, like the skin of a real drum, and it vibrates when sound hits it. This mechanical motion is passed

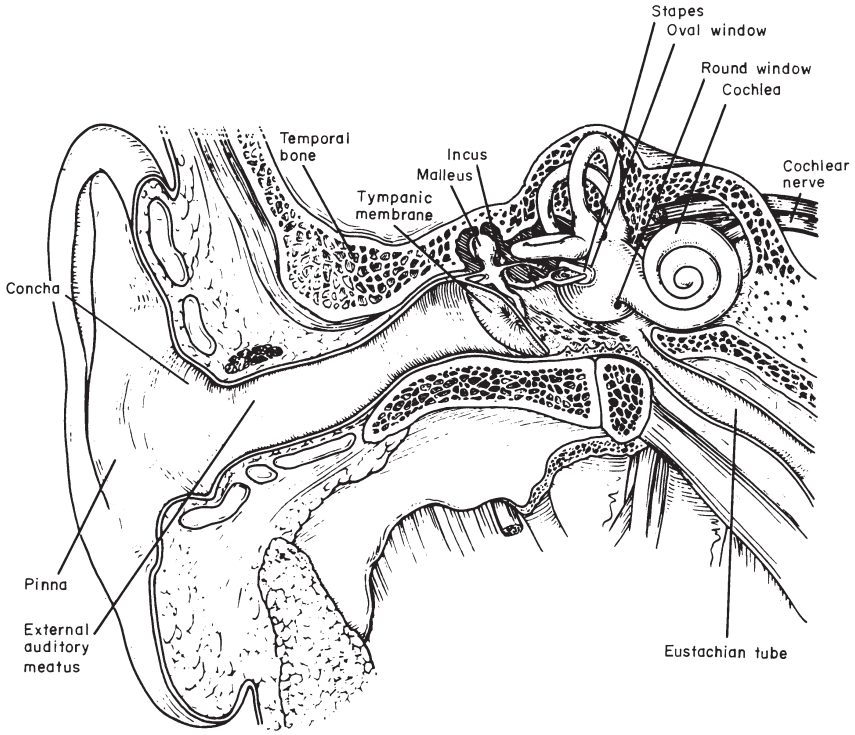


Figure 1.1: Outer, middle, and inner ear in man. From [Pickles, 1988].

down to the ossicles (malleus, incus, and stapes), which make up the middle ear. The function of the middle ear is to pass the mechanical vibration to the inner ear, counter balancing the difference in impedance between the air in the outer ear and the fluid of the inner ear. The combined effect of both the outer and the middle ear can be modeled by a linear filter depending on the direction of the sound.

The cochlea or inner ear is shown in Figure 1.2. The cochlea performs the most important transformation of the sound before it is transmitted to the brain as a neural code. The cochlea resembles a spiraled tube filled with fluid and divided by two membranes into three compartments or scalae. The membrane between scala media (the middle compartment) and scala vestibuli (the upper compartment) is called Reissner's membrane. The membrane between scala media and scala tympani is called the basilar membrane. The organ of Corti is located on top of the basilar membrane and comprises the sound receptors or hair-cells. The hair-cells are named after the thin hairs, called stereocillia, sticking out of them. The hair cells are connected at their bases to the 30000 nerve fibers of the auditory nerve.

The vibration of the stapes, connected to the oval window of the cochlea, elicits movement in the fluid of the cochlea. This pressure wave inside the fluid causes the basilar membrane to move up and down, starting at the base near the oval window and moving further to the apex or top of the cochlea. Because the basilar membrane

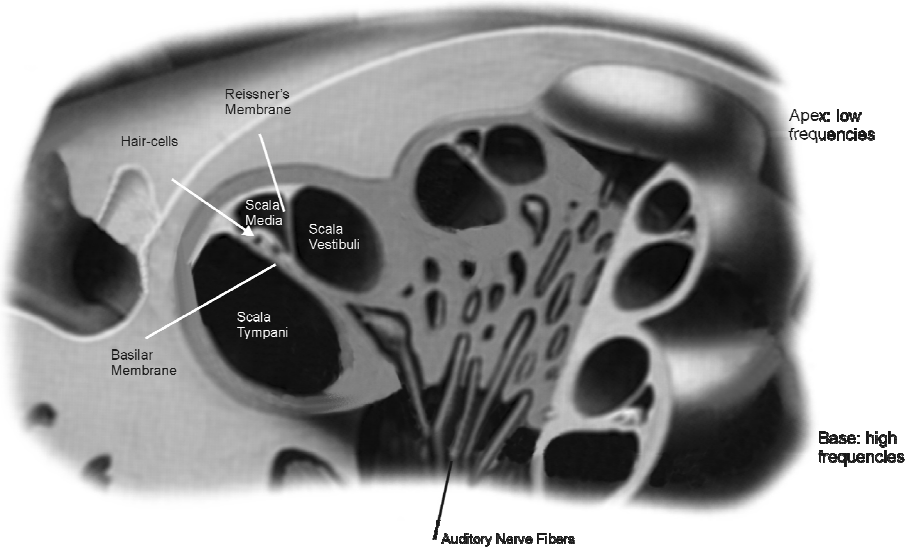


Figure 1.2: The inner ear or cochlea. Artist impression of the cochlea. The cochlea consists of a stapered and spiraled tube making three turns in total. In this drawing a part of the lower two turns is cut-away to show the internal parts of the cochlea.

is narrower and stiffer near the base, the fast or high frequency vibrations are mostly located near the base and lower frequency vibrations are most prominent near the apex. This important mechanism which separates high from low frequencies along the length of the cochlea is called tonotopy. The cochlea can be modeled as a filter bank, where every point along the basilar membrane behaves as a different filter from the filter bank. Each filter can be modeled by a gammatone filter [Slaney, 1993] with center frequencies spaced approximately logarithmic from apex to base [Greenwood, 1990] and the bandwidth of the filters also increases with distance along the basilar membrane from apex to base [Glasberg and Moore, 1990].

The moving up and down of the basilar membrane causes the hairs of the hair cells to bend, starting a complex mechanism of ion-channels and depolarisation. This depolarisation in turn commences the release of neurotransmitter towards the auditory nerve. The auditory nerve reacts to the neurotransmitter by generating action potentials or neural spikes which transfer the information about the incoming sound to the brain. The higher the intensity of the sound, the more neural spikes are sent to the brain.

In persons suffering from a hearing loss, the above outlined functioning of the ear is impaired in some way. When the outer or middle ear is impaired, the hearing loss is in general not severe and can be remedied with a classical hearing aid or through surgery. A classical hearing aid can also be an effective remedy when there is only mild damage to the inner ear. However, for most cases of severe to profound hearing loss there is significant damage to the inner ear with a large number of hair cells dead or absent. In this case only a cochlear implant can provide some rehabilitation.

1.1.3 History of cochlear implants

The Italian scientist Allesandro Volta is credited with being the first to demonstrate that electrical stimulation could directly evoke auditory sensations in humans. When he invented the battery around 1800, some 200 years ago, he placed the two sides of a 50-volt battery to his ears. He then observed "une secousse dans la tête" (a shock within the head), and some moments after "I began to hear a sound, or rather a noise in the ears, which I cannot well define: it was a kind of crackling with shocks, as if some paste or tenacious matter had been boiling (…) The disagreeable sensation, which I believe might be dangerous because of the shock in the brain, prevented me from repeating this experiment."

After Volta's unpleasant experience it lasted almost another 150 years before the interaction between electric stimulation and hearing was further investigated. In the 1930's and 1940's Stevens and colleagues showed that there was a mechanism that enabled to excite the auditory nerve directly through electrical stimulation with electrodes placed inside the ear canal [Stevens, 1937; Stevens and Jones, 1939; Jones et al., 1940].

It were however two French scientists who provided the first detailed description of a cochlear implant in a deafened person [Djourno and Eyries, 1957]. This patient described the sound as coming from a cricket ("cri du grillon") when the level was low, and like a whistle ("sifflet à roulette") when the level was higher. Moreover, this patient was almost immediately able to recognize the words father and mother ("papa" and "maman") and learned progressively to recognize more words.

Similar work was later done in the 1960's and the 1970's by scientists at the west coast of the United States of America where, among others, the first multi-channel implant was designed and implanted first in animals and later in man. After being in a research state for more than 20 years, in 1984 the first commercial cochlear implant (the House-3M single-electrode implant) was approved by the Food and Drug Administration of the U.S. This marked the beginning of the wide spread of cochlear implantation for deaf subjects over the entire world. Several universities developed cochlear implant systems, for instance the Flemish Laura implant that was developed at the University of Antwerp in Belgium and the Ineraid or Symbion implant which was developed at the University of Utah in the U.S.

At present, almost all cochlear implants are manufactured by one of three companies: the Clarion implants from Advanced Bionics Corporation (U.S.), the Pulsar implant from Med-El (Austria), and the Nucleus implants from Cochlear Ltd. (Australia).

1.1.4 The cochlear implant device

Every implant that is currently implanted consists of two parts: the internal part which is implanted inside the patient's head, and the external part which is worn by the patient. The external part consists of a Behind-The-Ear (BTE) sound processor and a transmitting coil which are linked with a small cable (see Figure 1.3.A). The BTE contains the microphone, the batteries, a digital sound processor, some buttons to change the program or to change the volume, and possibly a LCD display. The

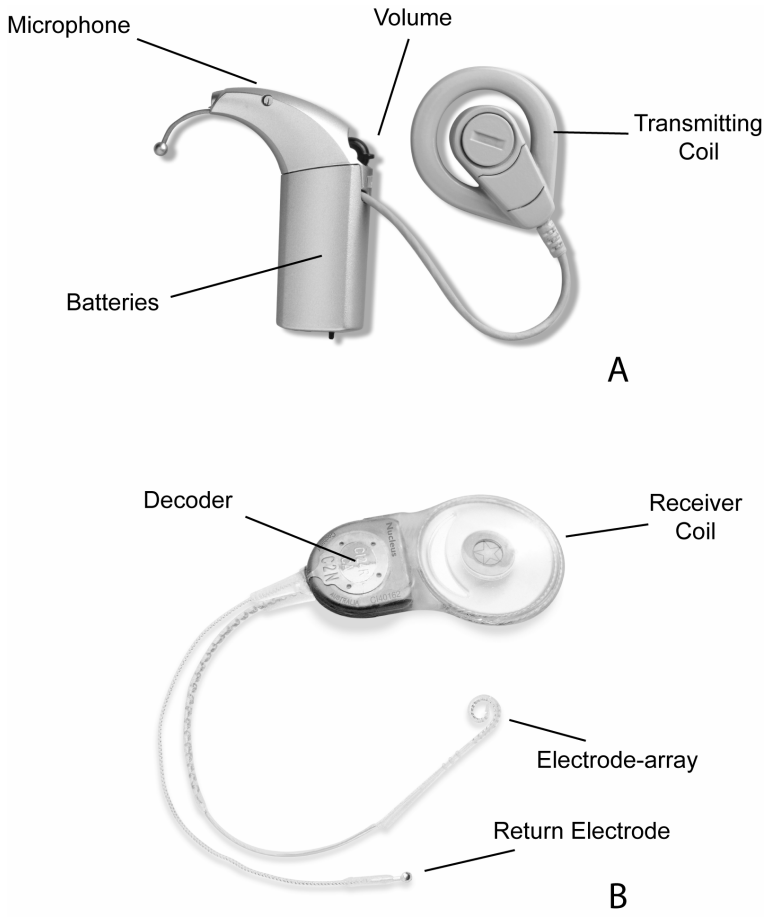


Figure 1.3: The parts of a cochlear implant. Panel A shows the external part of the cochlear implant. Panel B shows the internal part of the cochlear implant. [Cochlear Ltd].

internal part consists of a receiving coil or antenna, a decoder and stimulator, and the electrode array.

The electrode array consists of two silicon carriers: one carrier is inserted into the cochlea and contains up to 22 electrode contacts, the other carrier is implanted beneath the skin of the patient and contains one ball electrode which can serve as a return electrode (See Figure 1.3.B). Older devices also had a body-worn speech processor where the calculation of the current amplitudes based upon the incoming sound (see 1.1.5) was done, but advances in power-consumption and miniaturization have enabled the engineers to incorporate all parts in the BTE.

The surgery for implanting the internal device starts by exposing the mastoid by raising a “flap” of soft tissue (skin and muscles) from behind the ear (see Figure 1.4 panels A through C). Intraoperative flap design and plans for device positioning are aided by use of a mock implant and a mock behind-the-ear processor. Then a simple

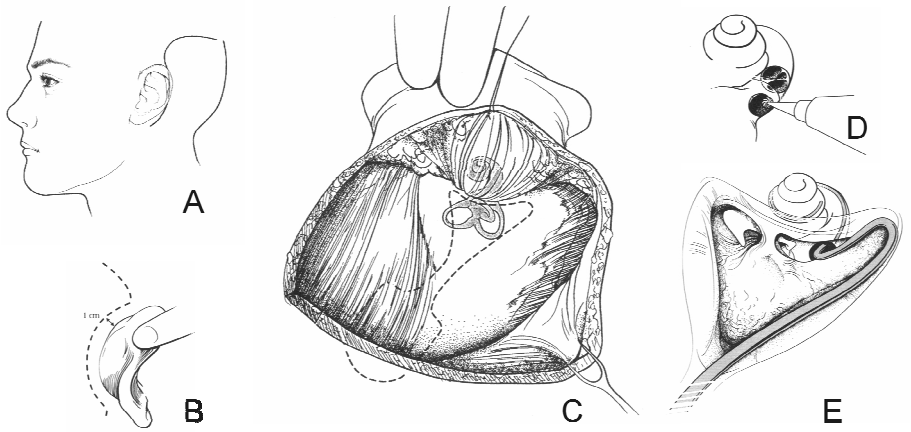


Figure 1.4: Steps in the implantation of a cochlear implant. Panels A and B: The line for incision to create the flap. Panel C: The flap of skin and muscles raised to access the mastoid. The dashed line indicates the position of the well where the receiver of the implant will be placed. Panel D: Drilling of the cochleostomy (opening of the cochlea) into the scala tympani. Panel E: Insertion of the electrode array into the opened cochlea. Adapted from [Niparko, 2004].

mastoidectomy is performed by drilling it away with a miniature rotating drill like that of a dentist. The resulting bony ledge can serve as a well for positioning and fixing the receiver coil of the internal part (See Figure 1.4.C). Subsequently, the middle ear is opened through the facial recess approach. In this way the surgeon can look directly to the round window of the cochlea. The cochlea is opened just below the basilar membrane so that the scala tympani is accessible (Figure 1.4.D). The electrode array is inserted into the scala tympani of the cochlea under direct visualization, minimizing trauma to the different membranes of the cochlea. Full insertion of the array requires an insertion depth of 25 to 30 mm (See Figure 1.4.E). After insertion the receiver is fixed to the skull of the patient and the opening is closed with sutures. The device is switched on for the fist time approximately one month after surgery.

When the device is turned on, the microphone(s) inside the BTE will pick up the acoustic wave and transform it into an electrical signal. This signal is then digitized and processed in the digital sound processor. This digital sound processor will calculate all the currents that have to be delivered on each electrode depending on the incoming sound. This information is then encoded and sent through the external transmitter-coil on a radio frequency between 2 and 10 MHz. The internal receiver-coil will pick up this signal and decodes it so that the current source(s) in the stimulator apply the correct amount of current on each electrode.

The electrode contacts implanted in the cochlea are spaced equidistantly from base to apex (See Figure 1.5). When current is injected through one of these electrodes, the auditory nerves close to that specific electrode will start to react and send action potentials to the brain. When high frequency sounds are processed by the cochlear implant, the electrodes close to the base of the cochlea will be stimulated because in

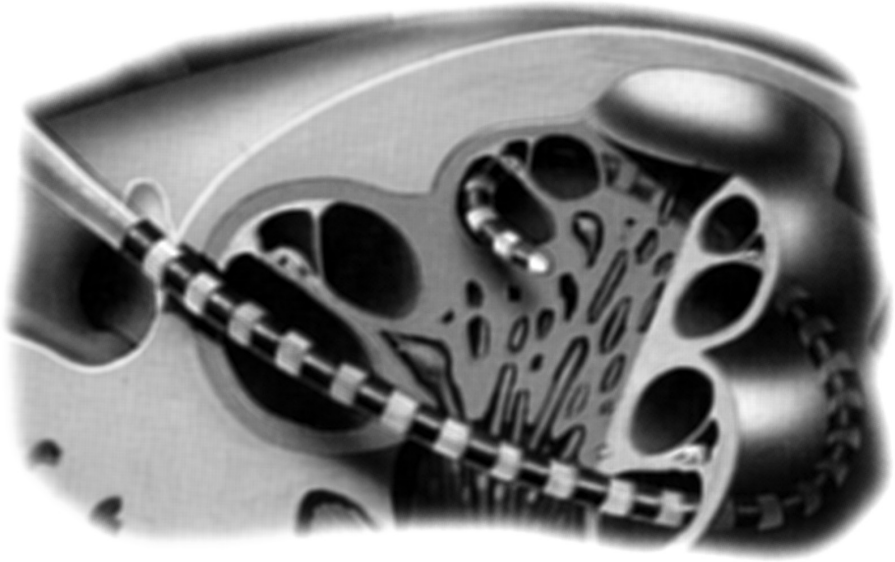


Figure 1.5: Detail of cochlear implant inserted into the cochlea.

a normal ear that region would also be responsive. When low frequency sounds are processed, the electrodes located more towards the base are stimulated.

There is however a large problem in separating the high frequency sounds from the low frequency sounds through a cochlear implant because there is a very large overlap between excited populations of neighboring electrodes. This phenomenon is called channel interaction and is caused by the current spread in the cochlea. When current is injected into the cochlea, this current does not flow directly towards the desired nerves. Instead, the current spreads out over almost the entire cochlea decaying in intensity from the point of injection. This is because the fluids inside the cochlea are relatively good conductors that easily pass on the electrical current.

1.1.5 Sound processing in cochlear implants

Throughout the past decades that cochlear implants already have existed, a large number of speech processing strategies have been proposed (for an overview see [Wilson, 2000]). Currently only a limited set of processing schemes, providing the best speech perception, still remain and are implemented in commercial cochlear implant sound processors. These strategies can roughly be divided into two classes: sound processing strategies with discrete interleaved stimulation, and sound processing strategies with continuous stimulation. In the former group the different channels are stimulated interleaved while in the latter group all, or at least part of the channels, are stimulated simultaneously.

All of the current sound processing strategies of the first group, with interleaved pulses on the different channels, are derived from the CIS (Continuous Interleaved Sampling) strategy proposed by Wilson and colleagues [Wilson et al., 1991]. A block diagram of the CIS strategy is shown in Figure 1.6. This CIS strategy has become a standard in cochlear implant sound processing and is currently

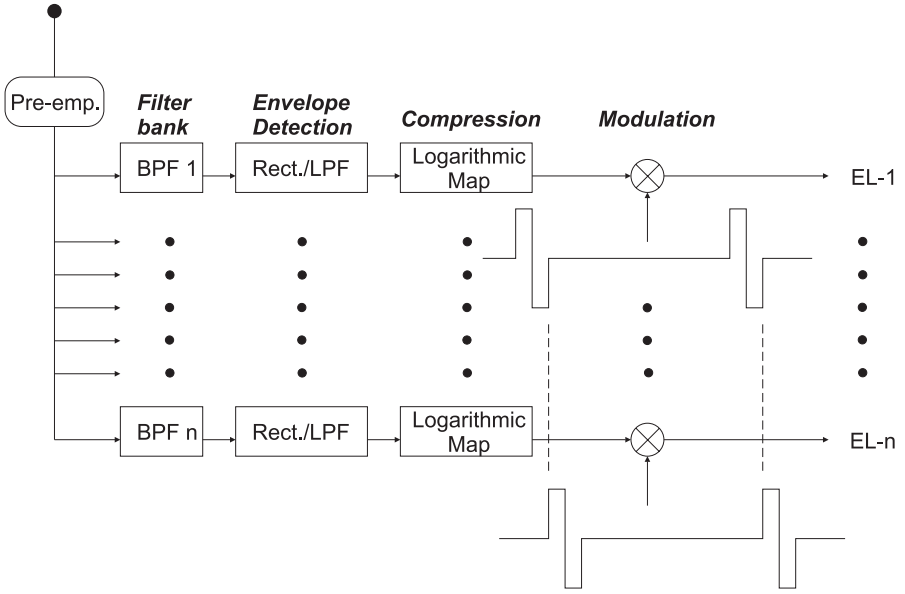


Figure 1.6: Block-diagram of signal processing in CIS scheme. Adapted from [Wilson, 2004].

implemented in the sound processors of each major cochlear implant manufacturer. Inputs from a microphone and automatic gain control (AGC) are directed to a pre-emphasis filter, which attenuates frequency components below 1.2 kHz at 6 dB/octave. Roughly speaking, this pre-emphasis filter flattens the average spectrum of speech and somewhat mimics the functions of the outer and middle ear.

The output of the pre-emphasis filter is directed to a bank of bandpass filters. This filter bank splits up the incoming audio channel in N different channels by passing each part of the spectrum to only one channel using band-pass filters (the number of channels N most often corresponds to the number of active electrodes in the implant). This function is similar to the physiology of the cochlea where high frequencies are separated from low frequencies along the length of the basilar membrane. The cut-off frequencies of the filters are usually spaced nearly logarithmically based upon the spacing of center frequencies along the basilar membrane [Greenwood, 1990] or based upon the spacing of psychoacoustically obtained filter shapes of the normal ear (ERB or Bark scale) [Glasberg and Moore, 1990]. The total spectrum of the filter bank usually spans from around 150 Hz up to 8 kHz. After the filter bank, each channel is followed by an envelope detector. This is usually accomplished by rectification and low-pass filtering (with cut-off frequencies typically between 200 and 400 Hz) as these mimic the physiology of the hair cells. However, a Hilbert transform [Hilbert, 1912] is also often used as an envelope detector because this may improve the computational cost in the digital sound processor (see for instance the ACE strategy discussed below and the F0mod strategy proposed in this thesis). In order to obtain the envelope based upon the Hilbert transform, the analytic signal $s(t) = s_r(t) + \sqrt{-1}s_i(t)$ is calculated for each

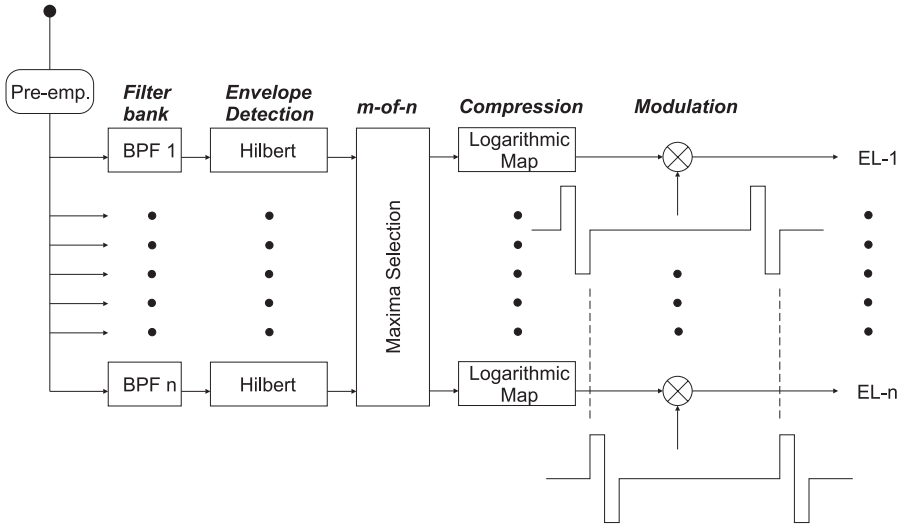


Figure 1.7: Block-Diagram of signal processing in ACE scheme.

channel, where $s_r(t)$ is the filter output in one channel, $s_i(t)$ is the Hilbert transform of $s_r(t)$. The Hilbert envelope is the magnitude of the analytic signal $A(t) = |s(t)| = \sqrt{s_r(t)^2 + s_i(t)^2}$ [Ville, 1948]. In practice, the Hilbert transform is often combined with the band-pass filtering using complex filters whose real and imaginary parts are in quadrature, such as in a FFT.

A logarithmic transformation is used to map the relatively wide dynamic range of derived envelope signals (in the order of 30 dB, but some processors allow up to 80 dB) onto the narrow dynamic range of electrically evoked hearing (in the order of 2 to 15 dB, depending on the subject and stimulus conditions). The dynamic range is defined by the difference from the T-level to the C-level. The T-level corresponds to the minimal electrical current that is required to elicit an auditory sensation in the particular patient and for the particular channel. The C-level corresponds to the level of the electrical current required to elicit a relatively loud but still comfortable auditory sensation. A second function of the compression is to remove the difference in loudness growth between acoustic hearing and electric hearing. In acoustic hearing loudness is a compressive function of intensity, while in an electric hearing loudness grows very rapidly (exponentially) with increasing current intensity [Zeng and Shannon, 1994].

The compressed channel outputs are used to modulate trains of biphasic pulses. The modulated pulse trains are applied through the RF transmission link to the corresponding electrodes in the cochlea. Pulse trains derived from low-frequency channels are directed to apical electrodes in the implant, and pulse trains derived from high-frequency channels are directed to basal electrodes in the implant.

A key feature of the CIS strategy [Wilson et al., 1991], and related strategies, is that none of the various pulses on the channels overlap in time, but that they are interleaved. This eliminates a major component of interaction among channels that

otherwise would be produced through vector summation of the electrical fields from different (simultaneously stimulated) electrodes. Such interaction would be expected to reduce the salience of channel-related cues.

A number of variations exist to the CIS strategy. These include the Speak (Spectral Peak; [Seligman and McDermott, 1995]), ACE (Advanced Combination Encoder; [Vandali et al., 2000]), and other "m-of-n" strategies [Ziese et al., 2000]. The main difference of these strategies with the CIS strategy is that not all electrodes are stimulated on a sweep over the electrodes but only a selected number of them. The main advantage of such a reduction of stimulated electrodes is that the stimulation rate on the selected channels can be raised and thus more temporal accuracy is available on these selected channels.

The functional block overview of the ACE strategy is given in Figure 1.7. The ACE scheme is the default sound processing strategy for the implant system which is implanted the most internationally, the Nucleus CI24 implant. In comparison to the CIS scheme an extra processing block is inserted after envelope extraction. This block performs maxima selection, i.e. only the m most intense channels are selected and the other channels are put to zero. In this way only m (where m varies between 6 and 12) out of the 22 channels are stimulated on each processing cycle. Because the maximal stimulation rate of the Nucleus CI24 implant is 14400 pulses per second, only stimulating a selected number of the 22 electrodes allows stimulating at rates up to 2400 pulses per second and per electrode. In contrast, if no maxima selection would be performed, as in the CIS strategy, only 650 pulses could be stimulated per second and per electrode. This latter rate might be too low for optimal transmission of the temporal details of speech because the pulse rate should be about 4 to 5 times higher than the highest modulation rate in cochlear implants [McKay et al., 1994].

The second group of cochlear implant sound processors do not interleave their pulses like the CIS strategy (or its variants). Most of these current strategies are derived from the CA strategy [Eddington, 1980]. Because a great advantage was found for CIS-like strategies over the CA strategy [Wilson et al., 1991], most recent strategies are derived from the CIS strategy. However, the SAS strategy implemented in the Clarion device from Advanced Bionics uses simultaneous stimulation like the CA strategy, but also incorporates a number of features to minimize channel interaction [Battmer et al., 1997]. With this improved simultaneously stimulating strategy performance was reported to be comparable to the CIS strategy [Osberger and Fisher, 2000].

More recently, an intermediate strategy, denoted HiRes, was implemented in the Clarion CII Device [Filipo et al., 2004; Koch et al., 2004]. This strategy is similar to the CIS strategy in that it interleaves the pulses on the different channels. It, however, also resembles the CA and SAS strategies in that it does not include envelope detection but rather the compressed filter output is used to modulate the pulse trains. The initial results indicate a clear benefit in speech perception over the older sound processing strategies like the CIS strategy. However, the comparison strategies in those studies were limited to 8 channels (16 channels in the HiRes strategy) and to relatively low pulse rates: 833 relative to more than 3000 pulses per

second per channel in the HiRes. These differences might also have biased the results of those studies.

1.1.6 Outcomes of cochlear implantation

Although cochlear implantation clearly entails inherent risks associated with the surgery, post-surgical complications are limited. About 2.3% major interventions are necessary during follow-up [Bhatia et al., 2004]. The decades of implant experience have also improved the reliability of the implanted devices, although manufacturers are scarce with information about implant breakdown. For adults with the older Nucleus CI22M implant 5 years after implantation 98.4% of the devices were still functioning normally and after 15 years 96.6% was still functioning normally. For adults with the newer CI24M implant 99.6% of the devices were still functioning 5 years after implantation [Cochlear Ltd., 2003]. In children the implant appears to fail more rapidly, possibly because children fall more often on their head.

Clinical observations in patients with current sound processors indicate that for patients with more than 6 months of implant experience, the mean score on open-set word recognition varies between 25% and 40%. However, subjects vary strongly in their performances among each other as their scores go from 0% up to 100%. Results achieved with the most recently developed speech processing strategies reveal scores above 75% correct on words-in-sentence recognition, again with a range of 0% to 100%. These scores have significantly improved as speech processing strategies have evolved. A number of recipients are even able to communicate easily over the telephone. When the duration of deafness is shorter and the experience with their implant becomes larger, the performance of the subjects typically becomes better.

Capabilities for understanding speech in noise can be assessed by varying the noise level relative to the level of the signal (or the signal-to-noise ratio). Normal-hearing individuals can still comprehend sentences when the noise is about 5 dB louder than the speech. Implant recipients typically show degraded speech recognition when signal-to-noise ratios are lowered and generally fail to understand speech when the speech is less then 5 dB louder than the noise. There is thus a difference of 10 dB between the performance of CI users and normal-hearing persons.

Children with cochlear implants demonstrate significant improvements in closed-set speech discrimination, enhanced lip-reading, and most obtain some open-set speech understanding with their devices. Auditory skills appear to be higher when children are implanted at a younger age but little difference is seen when implanted before the age of one year compared to before the age of two [Svirsky, M.A., personal communication]. Children still improve their auditory skills years after implantation, and more than 75% of them reach substantial open-set recognition after 3 years of cochlear implantation [Niparko, 2004].

Deaf children generally develop slower than normal-hearing peers, especially at language related skills such as speaking, reading or writing. However, after receiving a cochlear implant an accelerated development is noticed. Unfortunately this accelerated development is for a lot of CI-children insufficient to completely bridge the gap with their normal-hearing peers [Svirsky et al., 2000].

Deaf children are almost never able to attend regular schools and seldom pursue higher education. In contrast, more and more deaf children with implants are starting to attend regular schools with some special help. In fact a number of implanted persons are studying at universities, including two of the test subjects that participated in the experiments of this thesis.

1.2 Pitch perception in normal-hearing persons

1.2.1 What is pitch?

Pitch is defined as *"that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low"* [American National Standards Institute, 1994]. An important aspect of this definition is the term *"sensation"*. This term means that pitch is a subjective quality dependent on human perception, and that it can not be measured objectively with any instrument. In order to measure and investigate pitch we are bound to subjective human reports about their sensations, which can possibly be biased or ambiguous. The fact that pitch is a subjective quality thus makes pitch research relatively complex. Somewhat stricter definitions of pitch also state that variations in pitch give rise to a sense of melody [Plack and Oxenham, 2005a].

Table 1.1: Frequencies of notes in tempered scale.

Note	Note name	Frequency (Hz)
A3	la	220.00
A3#	la #	233.08
B3	si	246.94
C4	do	261.63
C4#	do #	277.18
D4	re	293.66
D4#	re #	311.13
E4	mi	329.63
F4	fa	349.23
F4#	fa #	369.99
G4	sol	392.00
G4#	sol #	415.30
A4	la	440.00

It has been shown that, in general, pitch is related to the repetition rate of the waveform of the sound. As discussed further, this corresponds to the frequency of a pure tone, and to the fundamental frequency of a harmonic complex. This frequency is generally expressed in Hz (cycles per second). Table 1.1 shows the frequencies corresponding to a number of musical pitch heights or notes. Each note corresponds to exactly one specific frequency (for instance A4 corresponds to 440 Hz) and higher notes correspond to higher frequencies. When a note is raised by one octave, the frequency is exactly doubled. For instance, going from A3 to A4 corresponds to changing the frequency from 220 Hz to 440 Hz. Each octave is divided into twelve



Figure 1.8: Mandarin Chinese symbols for 'mother' and 'horse' on the left and right hand side, respectively. Both words are pronounced 'ma' but with different pitch contours.

steps, called semitones, and this is the distance between two successive notes or tones. When two notes are separated by one semitone, the frequency of the higher note is 5.95% higher. A semitone is then divided into 100 equal parts denoted cents. A separation between two notes of two semitones is called a *'second'* and a separation of 7 semitones (approximately 50% frequency difference) is called a *'(major) fifth'*.

Most probably, pitch is the most important dimension in music. Pitch is the basis for other musical concepts as key, consonance, chords, harmony, contour, etc. The melody of a song is carried by the variations in pitch of the notes played by the instruments. When these pitch variations are dropped, only the rhythm remains and music becomes rather dull. Thus, pitch is crucial for music perception.

Although pitch is not essential for speech understanding of Western languages, i.e. Western speech can easily be understood when the pitch contour is changed, pitch does carry suprasegmental information. This means that pitch and pitch changes provide information about the intonation (whether a sentence is a statement or a question), the emotion(s) of the speaker, the identity of the speaker, or his or her dialect. Pitch can also be used to highlight a word in a sentence or put stress on a specific part of a sentence.

Pitch has, however, different effects in different languages. In Western languages pitch carries only supplementary information and this is in clear contrast to tonal languages where pitch does have the ability to change the meaning of a word or phrase. In tonal languages, the pitch pattern is characteristic for the meaning or the grammatical function of the words. This means that words can be pronounced with different pitch contours, called tonal patterns, and these different pitch contours lead to different meanings. Tonal languages, such as Mandarin Chinese, Cantonese, Thai, Igala, etcetera, are mostly spoken in South-East Asia and Africa, however they are spoken by a very large population. Over 1 billion speakers use Mandarin Chinese. An example of a Mandarin Chinese word with the meaning changing according to the tonal pattern is given Figure 1.8. The word *'ma'* pronounced with a flat tonal contour, i.e. the pitch does not change during the word, means 'mother', and is shown on the left hand side of Figure 1.8. However, when the word *'ma'* is pronounced with a falling-and-rising tonal pattern, the meaning is changed into 'horse' (shown on the right hand side of Figure 1.8). For foreigners, learning to speak Mandarin Chinese, this can lead to unpleasant misunderstandings.

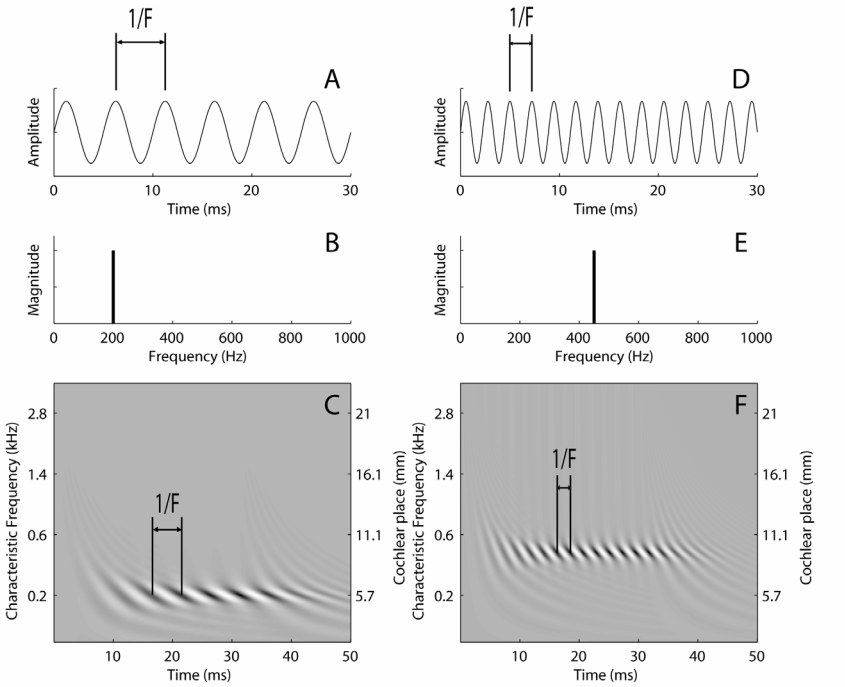


Figure 1.9: Pure tones or sinusoidal signals. The panels in the left, and right hand side columns show a pure tone at 200 Hz and 450 Hz respectively. The upper panels show the waveforms of both pure tones. The middle panels B and E show the spectrum of the respective pure tones. The bottom panels show the neural excitation from these pure tones as obtained from the AIM model (expressed as spikes per second) [Patterson et al., 1995].

Apart from using the pitch percept per se to rank stimuli from high to low, human listeners also use pitch information to segregate multiple simultaneous sound sources [Darwin and Carlyon, 1995]. This allows for example, combining the different components of a single vowel so that it can be identified. Secondly, pitch information also allows separating two simultaneous speakers with different pitches (for instance a male and a female). This allows normal-hearing subjects to focus attention on one voice and better understand speech in very noisy environments.

1.2.2 Pure tones and harmonic signals

One of the most simple and probably the most used sounds in auditory research is the pure tone or sine. A pure tone is completely defined by its amplitude and frequency. In Figure 1.9 the waveforms of two pure tones are shown. The pure tone in the left column has a frequency of 200 Hz, while the frequency of the pure tone on the right hand side is 450 Hz and the second pure tone thus sounds higher in pitch. Both pure tones have equal amplitude.

In the bottom panels of Figure 1.9 the neural excitation pattern in the auditory nerve is shown for the two pure tones respectively. The neural excitation is stronger in

neurons which have characteristic frequencies close to the frequency of the pure tone. Further away from this place along the basilar membrane the neural excitation diminishes (moving along the Y-axis in the figure). The neural excitation is also repetitive with the same frequency as the pure tone because of the phase locking. In other words, the neural excitation in the auditory nerve is also large for every 'top' of the pure tone, and is weaker for every 'dip' of the pure tone.

When the neural excitation patterns of both tones are compared, two differences are apparent:

1. The neural excitation of the higher frequency pure tone is located more towards the base. This is a spatial component.
2. The fluctuations in the neural excitation pattern are faster for the higher frequency pure tone. This is a temporal component.

Pure tones elicit very clear and highly salient pitch sensations in normal-hearing subjects and these sensations are related to the pure tone's frequency. Higher frequencies are perceived with higher pitches and the amplitude of the pure tone (related to the loudness) does not affect the perceived pitch. Normal-hearing subjects are very good at discriminating pitch differences of pure tones and are able to tell two pure tones apart even for very small pitch differences of less than 0.2% or 5 cents (0.05 semitones). However, it should be noted that pure tones are almost never encountered in a natural environment.

Most natural sounds eliciting clear pitches in daily life are harmonic sounds. These include voice, singing, and most musical instruments such as violin, piano, guitar, flutes. Harmonic sounds consist of a sum of several pure tones, called harmonics. The frequency of each harmonic is an integer multiple of the fundamental frequency (F0) of the harmonic. Figure 1.10 shows the waveform, spectrum and neural excitation of a harmonic sound.

Harmonic sounds are periodic, this means that after every specific amount of time the same waveform appears (see Figure 1.10.A). This period of time after which the waveform repeats itself, is called the fundamental period and is equal to 1/F0. The spectrum of the harmonic sound consists of peaks at each harmonic frequency and the distance between these peaks is always equal to the fundamental frequency (F0).

Harmonic signals elicit pitches that correspond to their fundamental frequency, even though other components (harmonics) exist at higher frequencies. It is even so that harmonic signals elicit pitches at F0 when the amplitude of the fundamental is nil. In other words, a normal-hearing person listening to a harmonic sound without the fundamental frequency still hears the pitch at the fundamental frequency even though this frequency is physically not present in the sound. The phenomenon is called the pitch of the missing fundamental.

The pitches of harmonic signals are approximately equally salient as those of pure tones. Normal-hearing subjects can perceive pitch differences between harmonic sounds differing in fundamental frequency with less than 0.5% or 10 cents (0.1 semitones). Thus, between two adjacent keys on a piano keyboard, normal-hearing subjects are still able to discriminate 10 different pitches.

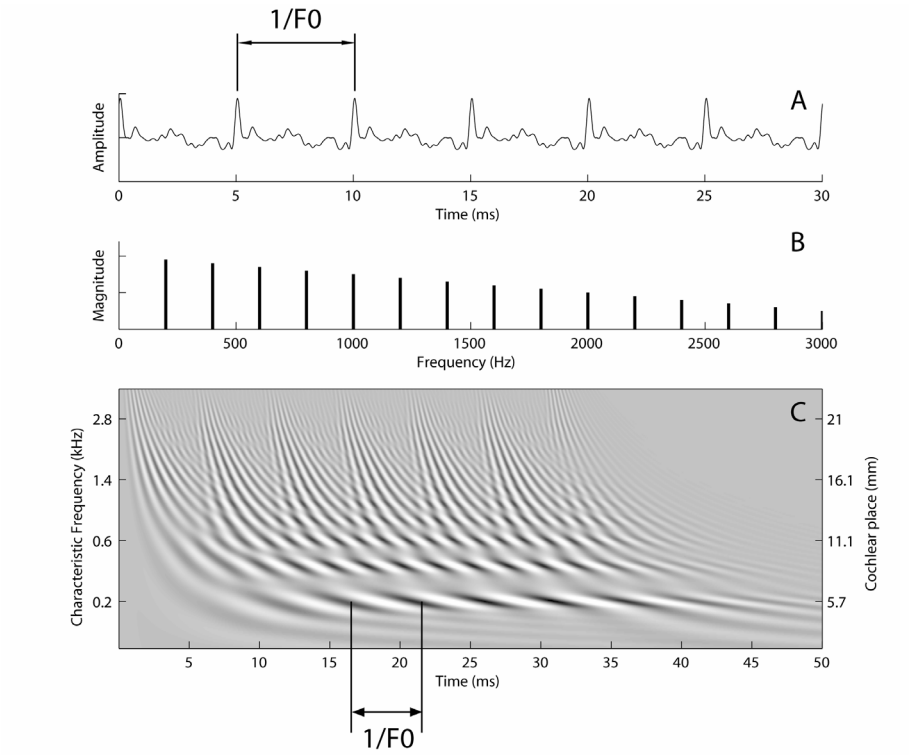


Figure 1.10: Harmonic signal. The upper panel shows the waveform of a harmonic signal. The middle panel B shows the spectrum of the harmonic signal. The bottom panel shows the neural excitation from this harmonic signal. The neural excitation was calculated by the AIM model [Patterson et al., 1995].

1.2.3 Pitch mechanisms

The first theories on pitch perception already existed 600 years B.C.: the string theory of Pythagoras. Newer, more general theories were developed over the following centuries by many researchers (see [de Cheveigné, 2005] for a review). However, a strong (and sometimes passionate) debate has always existed among researchers on how pitch is perceived and this debate still exists today. This section will review the state-of-the-art evidence on the mechanisms of pitch perception.

Currently, three different pitch mechanisms are revealed: spectral pitch, periodicity pitch, and purely temporal pitch. A diagram showing the three pitch mechanisms is shown in Figure 1.11. The existence of two different pitch mechanisms for spectral and periodicity pitch was already proposed 50 years ago by Licklider [Licklider, 1954]. The discovery of a third separate pitch mechanism for purely temporal pitch is more recent and still a matter of debate [Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Carlyon and Deeks, 2002; Grimault et al., 2002]. Evidence for the existence of the three separate mechanisms leading to pitch was obtained from different studies, including psychophysics, neurophysiology, and neuroimaging. Next we will discuss each pitch mechanism.

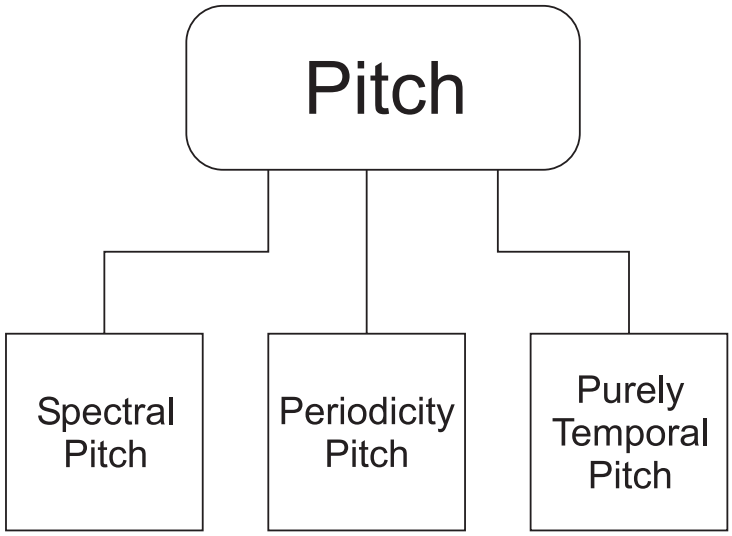


Figure 1.11: Pitch mechanisms in normal-hearing persons.

1. *Spectral pitch* is related to the overall spectral shape of the sound. For narrow-band sounds spectral pitch is determined by the centroid of the spectrum [Stover and Feth, 1983; Anantharaman et al., 1993; Dai et al., 1996]. For broad-band sounds spectral pitch is determined by the centroid of the excitation pattern [Zwicker and Fastl, 1999]. This means that for harmonic sounds the amplitude of the harmonics determines spectral pitch and not (at least to great extent) their frequencies. In other words, when the higher harmonics are more intense than the lower harmonics, the spectral pitch will be high. However, when only the lower harmonics are present the spectral pitch will be lower even though the fundamental frequency can be the same. Spectral pitch appears to be mediated by a place code, i.e. the location of the centroid of neural excitation along the cochlea determines the height of the spectral pitch. A number of researchers studying spectral pitch also called it sharpness of timbre, or brightness of timbre, because stimuli differing in spectral pitch can also be ranked from dull to sharp or from dim to bright.

2. *Periodicity pitch*, also denoted as complex pitch, is related to the fundamental frequency of a harmonic sound. This means that higher fundamental frequencies will lead to higher perceived pitches. The periodicity pitch mechanism is the most important pitch mechanism. It allows for the very good pitch discrimination found in normal-hearing subjects [Houtsma and Smurzynski, 1990] and it also elicits the highly salient pitch necessary for melody identification where purely temporal pitch [Moore and Rosen, 1979] or spectral pitch [Smith et al., 2002] fall short. Because periodicity pitch is only determined by the fundamental frequency of the sound it is a very robust percept. Reverberation or linear filtering of a sound, changing the amplitudes and phases of the harmonic components of the sound, do not affect periodicity pitch because the fundamental frequency remains constant. This also means that

different instruments can have the same periodicity pitch as long as their fundamental frequency is the same although their timbres clearly differ. The exact mechanism for periodicity pitch is not yet unravelled completely but it appears that both accurate place information and accurate temporal information are necessary in the peripheral neural representation [Oxenham et al., 2004]. This necessity for a high degree of resolution also has been interpreted as periodicity pitch relying on the *'fine structure'* of the stimulus and its neural excitation [Smith et al., 2002]. It should be emphasized that periodicity pitch relies on both place and temporal cues extracted from the neural excitation pattern, and not only on temporal cues as has been (and possibly still is) adhered by many researchers. However, it is obvious that without the correct place cues rate discrimination is limited [Carlyon and Deeks, 2002] and that normal-hearing subjects do not hear a single clear pitch for multiple harmonic frequencies [Oxenham et al., 2004].

3. On the other hand, it is clear that pitch differences can be perceived using solely temporal cues from the neural excitation pattern and that these pitch differences can even be used to transfer musical melodies [Burns and Viemeister, 1976]. This mechanism has been denoted *purely temporal pitch* or envelope pitch. Nevertheless, some researchers however opposed to purely temporal pitch as a pitch percept because some subjects denote changes in temporal pitch as changes in roughness or intermittency [Zwicker and Fastl, 1999]. Purely temporal pitch is determined by the rate of fluctuations in the envelope of the neural excitation. The higher this rate, the higher the perceived temporal pitch will be. The smallest rate difference that can be told apart by normal-hearing persons is larger than the smallest frequency difference they can perceive based upon periodicity pitch [Houtsma and Smurzynski, 1990]. Also purely temporal pitch appears to be limited by a central process so that changes in rates become less salient as the rates become higher. Above 300 to 600 Hz rates changes become imperceptible [Patterson et al., 1978; Carlyon and Deeks, 2002] while periodicity pitch works up to approximately 5000 Hz. Another important difference between periodicity pitch and purely temporal pitch is that the former mechanism appears to use all-order interval statistics of the neural spikes as in an autocorrelation [Meddis and Hewitt, 1991; Cariani and Delgutte, 1996]. While the latter mechanism appears to base its pitch estimate on a weighted sum of first-order intervals [Kaernbach and Demany, 1998; Carlyon et al., 2002].

The interaction and integration of the three different pitch mechanisms (spectral, periodicity, and purely temporal pitch) have not been completely investigated. It has been shown that the spectral pitch mechanism and its associated percept are independent from the two other mechanisms and their associated percepts [Demany and Semal, 1993; Plomp and Steeneken, 1971; Tong et al., 1983; McKay et al., 2000] and that spectral pitch is processed in a different region of the brain than periodicity and purely temporal pitch [Warren et al., 2003]. This means that a change in spectral pitch will not affect periodicity or temporal pitch, and vice versa. However, periodicity pitch and purely temporal pitch may not necessarily be independent. Very recent results indicate that the activation of one mechanism can

interfere with the other mechanism [Gockel et al., 2004] and both pitch mechanisms excite the same cortical areas in humans [Penagos et al., 2004].

The existence of multiple separate pitch mechanisms, as described above, means that pitch is *in se* a multi-dimensional percept. Apart from pure tones, where all pitch mechanisms vary consistently with frequency and thus the pitch is one-dimensional, all sounds evoke multiple pitch-like percepts simultaneously [de Cheveigné, 2005]. In a number of situations subjects are forced to combine these separate percepts into a single percept for instance to rank two sounds in pitch, or to indicate the height in pitch of a certain sound, or to identify the melody in a series of different sounds. In general, subjects use the most salient cue when they are forced to combine percepts which most often is periodicity pitch when its mechanism is activated. However, when none of the pitch cues of the different mechanisms are salient, subjects combine these percepts differently depending on condition and subject [Smoorenburg, 1970; Fearn and Wolfe, 2000; Zeng, 2002; Renken et al., 2004]. Also in chapter 4 we show that the integration of purely temporal and spectral pitch cues exists in normal-hearing subjects and the salience of the resulting combined cue is determined by the salience of the separate pitch mechanisms. That work also shows that when two pitch mechanisms give conflicting cues, the subject is unable to indicate a rise or a fall in pitch.

1.2.4 Measurement techniques for pitch

A large number of different techniques exist to study pitch perception. The most obvious technique is of course to query human subjects their percepts when presented with particular stimuli. When this is done in a controlled way with adequately designed stimuli, this technique is called psychophysics. This technique has been used most often to study pitch perception. Other studies have also been used to study pitch perception. The techniques in those studies are going from training animals in pitch perception tasks, over neurophysiologic experiments recording nerve responses (most often in animals), to neuro-imaging in humans using PET (Positron Emission Tomography), MEG (Magnetic EncephaloGram) and fMRI (functional Magnetic Resonance Imaging) scanning of the activation of the human brain.

In this thesis, only psychophysical techniques were used because these studies investigate pitch directly. More objective techniques such as neuro-imaging or physiology might show differences in neural activity but it might be hard to unambiguously relate them to pitch perception because pitch is a subjective quality, a sensation.

The two psychophysical techniques most often used for investigating pitch are pitch scaling and pitch discrimination. In pitch scaling the subject is presented with a sound and is asked to indicate the height of the sound (qualitatively or quantitatively) or equivalently the subject is presented with two sounds and the subject is asked to indicate the pitch difference or pitch interval between the two notes. The drawback of scaling techniques is that it may be hard to extract information on the sensitivity of the subjects to pitch changes.

In pitch discrimination, the subject is presented with two (or more) sounds and has to indicate the highest one in pitch. The difference between the two sounds is then made so small that the subject starts having difficulties in identifying which of the two stimuli sounds highest in pitch. This level is called the Just Noticeable Difference (JND) and is usually defined as the point where the subjects still consistently rank the stimuli in pitch 75% of the times (note that when two stimuli are presented the subject has 50% chance to have it correct just by guessing). Pitch discrimination experiments thus can provide accurate estimates of the sensitivity to pitch changes for a particular condition or subject. Because the main goal of this thesis is to assess and improve the pitch performance or pitch sensitivity in cochlear implantees most studies in this thesis are based upon pitch discrimination tasks.

In pitch discrimination, and psychophysics in general, the stimuli are presented a large number of times to the subjects and their responses are averaged over these different trials. The average proportion of times that the two sounds were correctly ranked in pitch is then calculated and this proportion (p) is a measure for the performance of the subject in the particular task. Very often this proportion is transformed into another metric called *d'* ("d-prime") based upon signal detection theory [Green and Swets, 1988]. For a *d'* of zero, the two presented stimuli are not discriminable from each other, for a *d'* value near 1 the stimuli are perceptually apart by their JND. The *d'* value can be regarded as a measure of the internal salience of the pitch percept normalized to its variance. The *d'* metric will often be used in the analysis of the experiments in this thesis.

1.3 Music and pitch perception in cochlear implant subjects

1.3.1 Music perception with a cochlear implant

Most research with cochlear implant recipients has focussed on speech perception as the primary goal of cochlear implants is to restore oral communication in deaf people. In comparison, the research on music perception in cochlear implants is limited. In the early days of cochlear implantation, reports on music perception were anecdotal or single case studies [Eddington et al., 1978; Moore and Rosen, 1979]. These results were very promising because they already indicated that melody recognition was possible, although performance was lower than for normal-hearing persons.

Studies investigating music perception in larger groups of cochlear implant subjects have indicated that among the subjects there exists a wide range of musical involvement and enjoyment [Gfeller et al., 2000]. Most subjects reported that music enjoyment is poorer than before their hearing loss. Further studies investigating the separate properties of music listening tried to identify the major impediments for music appraisal in cochlear implant subjects [Gfeller and Lansing, 1992; Gfeller and Lansing, 1991; Schulz and Kerber, 1993; Gfeller et al., 1997; Leal et al., 2003; Kong et al., 2004]. The general conclusion of those studies is that rhythmical structures of music are generally perceived (almost) as good by cochlear implant recipients as by

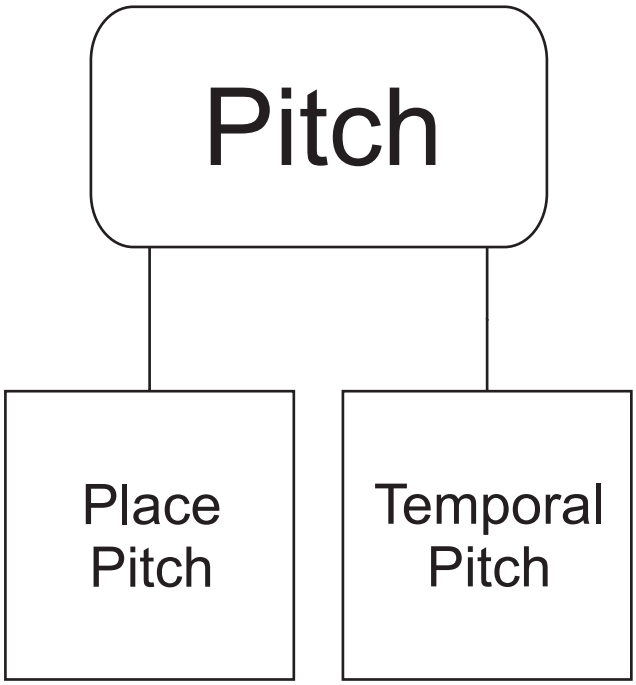


Figure 1.12: Pitch mechanisms in CI subjects.

their normal-hearing peers. However, cochlear implant subjects are markedly poorer in following the contour of a melody. This is for instance exemplified in that cochlear implant recipients generally prefer music with a clear rhythm and that they recognize more songs when the songs have a striking rhythm.

The largest problems for music perception with cochlear implants are in the perception of pitch patterns. For musical notes, the smallest difference in pitch that cochlear implant subjects can perceive extends over a range from 1 semitone up to 2 octaves (24 semitones) depending on the subject [Gfeller et al., 2002]. In comparison, normal-hearing subjects detect changes of less then one tenth of a semitone as mentioned earlier. A study comparing pitch identification performance with synthetic electrical stimuli and with musical notes processed through the subjects' speech processor, showed that the sound processing scheme limited significantly the subjects' performance [Pijl, 1997].

1.3.2 Mechanisms of pitch perception in cochlear implant recipients

In cochlear implant recipients two pitch mechanisms are observed: temporal pitch and place pitch [Eddington et al., 1978] (see Figure 1.12). The first mechanism (*temporal pitch*) is related to the temporal pattern of stimulation: as the frequency of stimulation (either pulsatile or sinusoidal) on a single electrode is increased, a higher pitch is perceived. The second mechanism (*place pitch*) is related to the spatial

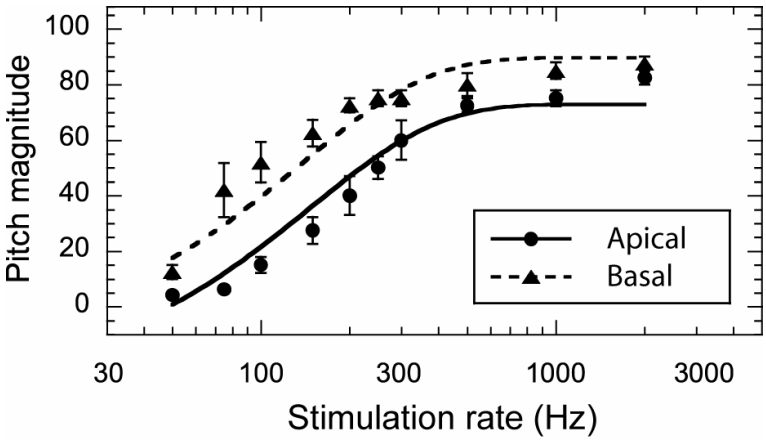


Figure 1.13: Temporal pitch estimates from a scaling experiment. Temporal pitch rises as the stimulation rate (on a single stimulation site) increases. Above 300 Hz, no further increase in pitch is noted. The pitch as measured from a basal electrode is higher as that of an apical electrode (triangles and dots) because of place pitch. Adopted from [Zeng, 2002].

pattern of stimulation: the more the stimulation is moved towards the base of the cochlea, the higher the perceived pitch. Both percepts are pitch like but not identical, and are interpreted independently of each other by the subjects [Tong et al., 1983; McKay et al., 2000]. Both temporal and place pitch are discussed in more detail in this section.

1. The first mechanism, temporal pitch or also rate pitch, is derived from the fluctuations in time of the neural excitation pattern. In electrical stimulation, as in acoustical stimulation, the auditory nerve fibres *phase lock* to the stimulation pattern, causing the neural excitation pattern to modulate at the stimulus frequency. In fact, the synchronisation of the spikes to the stimulus is better with electrical stimulation than with acoustical stimulation [Shepherd and Javel, 1997]. Consequently, cochlear implant subjects can use this cue in the neural excitation pattern to extract information about the frequency of the stimulus. When electrical stimulation is varied in frequency between 50 Hz and 300 Hz a constant rise in pitch is perceived by most cochlear implant subjects, which is called temporal pitch [Shannon, 1983a; Tong and Clark, 1985; Townshend et al., 1987; Zeng, 2002]. Below 50 to 75 Hz the repetition rate is not perceived pitch like but rather as an unpleasant rattling. When the stimulation frequency is increased above 300 Hz, the perceived pitch does not change any more (see Figure 1.13), unless for a few exceptions [Townshend et al., 1987].

   Discrimination experiments for temporal pitch, measuring the smallest difference in stimulation rate the cochlear implants can perceive, have lead to different results depending on the subject. However, a general trend is that the discrimination is best at low frequencies (below 200 Hz). For higher frequencies the smallest detectable frequency difference grows dramatically. This is consistent with the scaling data that show no difference in pitch above 300 Hz. Between 100 Hz and 200 Hz, the smallest noticeable difference varies between

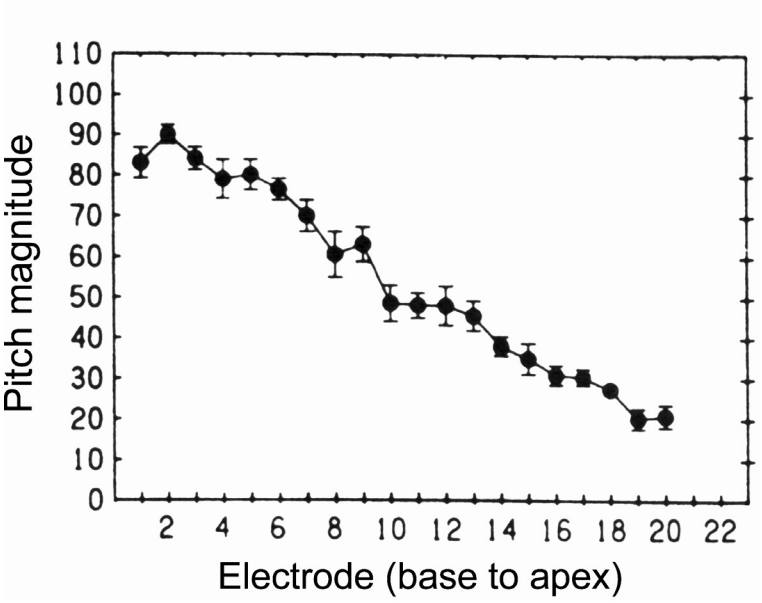


Figure 1.14: Place pitch estimates from a scaling experiment. As the stimulation site is changed from basal to apical electrodes, the perceived pitch falls. Adapted from [Cohen et al., 1996b]

3% and 40% depending on the subject (or between 0.5 and 6 semitones). The sensitivity based upon temporal pitch in CI subjects is slightly weaker than the sensitivity found in normal-hearing subjects based upon purely temporal pitch [Carlyon and Deeks, 2002]. This might be caused by the pathological state of the auditory nerve fibers of the implanted subjects or other factors such as the (long) duration of deprivation of auditory input.

Temporal pitch cues in cochlear implant recipients have been shown to provide a means for melody recognition [Pijl and Schwarz, 1995b] and musical interval identification [McDermott and McKay, 1997; Pijl, 1997].

2. The second pitch mechanism in CI subjects, place pitch, is derived from the place along the cochlea where the neural excitation is largest. This mechanism is enabled by the tonotopic organisation of auditory nerve fibers together with the feasability of restricted nerve excitation in the neighbourhood of the electrode. When an electrode located close to the base of the cochlea is stimulated, a restricted group of nerve fibers is activated that would be stimulated by high frequencies in a normal ear. This leads to a perceived high pitch. Similarly a more deeply inserted electrode (located more towards the apex of the cochlea) elicits low pitch percepts. This mechanism is called place pitch (see Figure 1.14). Sensitivity to changes in place pitch vary strongly across subjects and the smallest difference in shift of stimulation along the cochlea that subjects can perceive varies from 0.3 mm up to 8 mm [Nelson et al., 1995].

Apart from describing the sensation as going from high to low as the stimulating electrode is switched from base to apex, subjects sometimes describe the sensation as going from sharp to dull. This has lead to an ongoing debate among researchers whether place pitch is really a pitch, or rather an aspect of the sound's timbre, namely sharpness. The discussion is not settled yet and the conclusions obtained depend on the strictness of different definitions of pitch [Moore and Carlyon, 2005]. In the discussion of Chapter 7 of this thesis some arguments in favour of place pitch as a pitch percept are presented. More details on the place pitch mechanism in cochlear implant subjects can be found in Chapter 3.

Most studies investigating place or temporal pitch use stimuli that stimulate only a single channel of the cochlear implant. In contrast, in daily life cochlear implants stimulate all channels simultaneously (or temporally interleaved) as sounds are processed through the system. Very few studies report on pitch perception and pitch mechanisms of synthetic multi-channel stimuli and even less on pitch perception of multi-channel stimuli of processed acoustical sounds.

Very recently, a number of studies have assessed how the pitch of harmonic sounds is perceived by cochlear implant subjects when they are listening through their sound processors [Geurts and Wouters, 2001; Gfeller et al., 2002; Leal et al., 2003; Geurts and Wouters, 2004]. These studies measured the smallest difference in F0 that CI subjects could detect. Their results indicate that performance in pitch discrimination varies strongly across subjects and across conditions, and ranges from approximately 1 semitone up to more than 24 semitones with a mean threshold of approximately 7 semitones (50% frequency difference) [Gfeller et al., 2002]. It should be noted that this performance is very poor. In fact, normal-hearing persons perform on average about 30 times better on the pitch discrimination task than cochlear implantees.

The study of Geurts and Wouters [2001] showed that both temporal and place pitch mechanisms contribute to F0-discrimination in cochlear implantees. However, it was unclear how the separate place and temporal pitch cues were perceived and integrated. The performance varied very strongly over conditions and no clear reason for this variability was found. In Chapter 5 of this thesis the separate effect of place and temporal pitch cues on F0-discrimination is assessed. Also, the mechanisms of both place and temporal pitch with multi-channel stimuli are investigated (see Chapters 3, 5, and 7).

1.3.3 Relations between pitch mechanisms in NH and in CI

At present there exists a debate among researchers how the two pitch mechanisms in CI users (temporal pitch and place pitch) relate to the three pitch mechanisms in NH persons (spectral pitch, periodicity pitch and purely temporal pitch). This debate is closely related to more general debate whether normal-hearing subjects use rather spectral or temporal properties of the neural representation in the interpretation of a sound. A first group of researchers proposed the theory that normal-hearing listeners mainly use the positions of the individual components of a harmonic sound to group the components and estimate the F0 or pitch of the sound [Goldstein, 1973;

Terhardt, 1974; Cohen et al., 1995]. In this theory the major cue to the perceived pitch is thus the location or place of the excitation along the basilar membrane, therefore it could be inferred that periodicity pitch in normal-hearing listeners is related to the place pitch in CI subjects. In contrast, a second group of researchers proposed a theory stating that normal-hearing listeners use the timing information within single auditory nerve fibres to estimate the pitch of a sound [Meddis and Hewitt, 1991; Cariani and Delgutte, 1996]. From this second theory it could be inferred that periodicity pitch in normal-hearing listeners is related to temporal pitch in CI users. More recent theories on periodicity pitch perception however pose that periodicity pitch in normal-hearing persons is based upon both place and temporal cues in the neural excitation pattern [Shamma and Klein, 2000; Oxenham et al., 2004]. In summary, it is at present unclear how the mechanism for the periodicity pitch in normal-hearing listeners relates to any of the two pitch mechanisms in CI users.

For purely temporal pitch a relation can possibly be established between purely temporal pitch in normal-hearing listeners and temporal pitch in CI users. For instance, recent studies investigating the pitch of irregular rate pulse trains revealed similar results in CI users and in normal-hearing listeners when the pitch cues were either presented using the purely temporal pitch mechanism for NH listeners and using the temporal pitch mechanism for CI users [Carlyon et al., 2002; van Wieringen et al., 2003]. The results of these studies thus suggest an analogy between purely temporal pitch in NH listeners and temporal pitch in CI users.

In this thesis we investigate the possible relations between the two pitch mechanisms in CI users and the three pitch mechanisms in NH listeners. Based upon our results we state in chapter 8 that purely temporal pitch (NH) is closely related to temporal pitch (CI) and that spectral pitch (NH) is closely related to place pitch (CI). We found no evidence of an equivalent pitch mechanism for the periodicity pitch of normal-hearing listeners in the CI users.

1.4 Outline of thesis

1.4.1 Major research questions

In this thesis we deal with two kinds of research questions: fundamental psychophysical problems and more applied signal processing issues. Firstly, the psychophysical problems addressed in this thesis investigate the effect of the electrical stimulation patterns on perceived sound qualities, and more particularly pitch. The goal of these studies is to reveal the basic mechanisms in electrical (pitch) hearing. It is, to understand the physical and neural processes that elicit sensory percepts based upon an electrical stimulus. Secondly, the effect of the signal processing in the CI system on pitch perception is not known. The goal of the second kind of research in this thesis is to optimise the signal processing for accurate pitch perception because pitch is important for music and for understanding tonal languages. The design of this optimal sound processing for pitch perception will be based upon the findings of the psychophysical studies.

These two parts lead to the following research questions which will be addressed in this thesis:

1. How does the *place pitch* mechanism work when multiple channels are activated at the same time? What is the sensitivity in place pitch of broader multi-channel stimuli? To what physical attribute corresponds the pitch perceived from a multi-channel stimulus?
2. How do the different channels of a multi-channel stimulus interact with respect to the *temporal pitch* mechanism? Do the temporal pitch cues of different channels integrate constructively or do they interfere destructively? And does this integration or interference depend on physical properties of the stimulation pattern in each channel?
3. How are *temporal and place pitch* cues combined in a pitch ranking task? Can they be considered as completely independent or can one cue be affected by the presence of another (possibly conflicting) cue?
4. Is there some means or *model* by which we can study an analogy of the pitch percept in cochlear implantees through experiments with normal-hearing subjects? Can we elicit a pitch percept in normal-hearing subjects similar to what cochlear implantees hear? In other words, can we, normal-hearing persons, hear how a cochlear implant sounds?
5. How is *F0 related pitch information* encoded in the electrical stimulation pattern and which aspects of the sound processing affect this information? What is the effect of the type of *filter bank and envelope extraction* in the CI sound processing on F0 perception by cochlear implantees?
6. Is it possible and how is it possible to *optimise the current sound processing* schemes to improve pitch perception in cochlear implantees? And if this new sound processing does lead to improved pitch perception, does it also lead to improved music perception?

The first three questions are somewhat more fundamental and related to psychophysics. The answers to these questions will provide the basis for the further work and support the investigation of the next three questions which are more applied. The results of the studies measuring the effect of signal processing on pitch perception are fed back into the psychophysical knowledge. The overview of the major research questions is given in Figure 1.15.

All cochlear implant subjects that participated in the experiments were implanted with the Cochlear Nucleus CI24 implant. This implant was chosen as a platform for our research because, firstly, it had (and still has) the largest number of electrode contacts of all commercial implants on the market. Secondly, the presence of the European research and development division of Cochlear Ltd. (Cochlear Europe Technology Centre) in Mechelen allowed for a close collaboration with the manufacturer. And thirdly, we have access to subjects implanted in the U.Z. Leuven or subjects studying at the Katholieke Universiteit Leuven.

Participation of human subjects is ubiquitous to pitch perception research; fortunately a number of cochlear implant subjects is eager to devote time and energy

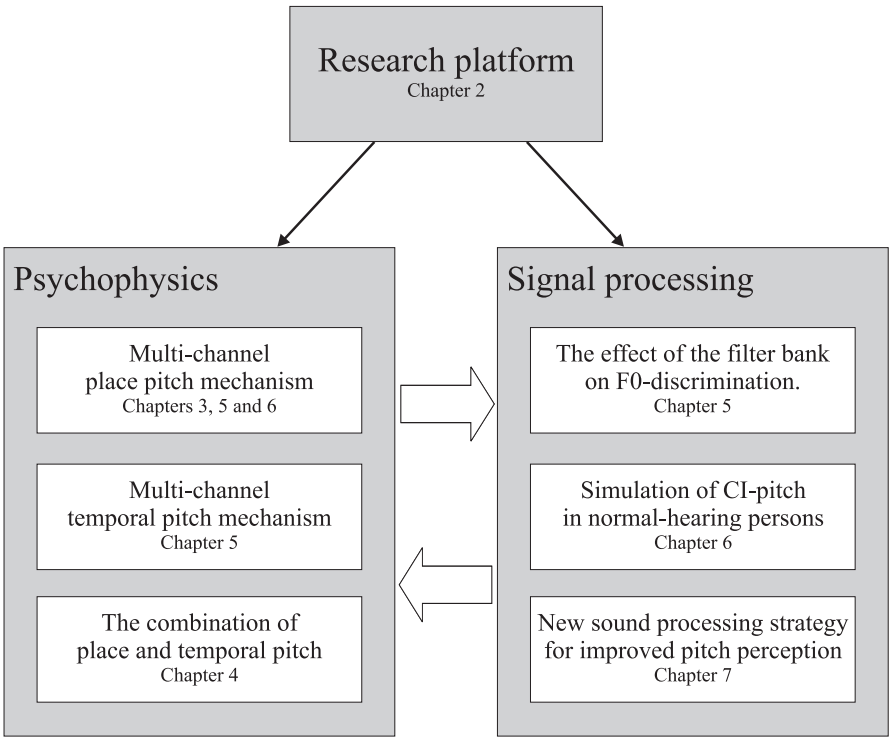


Figure 1.15: Schematic overview of the content of this thesis.

to numerous sessions of hours of repetitive tests. Though their enthousiasm was motivating, time available from each subject is not infinite.

1.4.2 Chapter by chapter overview

The research described in this thesis is divided into chapters where, in general, each chapter emphasizes one of the above stated research questions. The experiments described in chapters 3, 5 and 7 involve cochlear implant subjects. The experiments in chapters 4 and 6 involve normal-hearing persons with stimuli specifically processed to produce a model of CI.

Chapter 2 starts with the development of a research platform. Because pitch is a subjective percept it is necessary to do experiments with human subjects implanted with a cochlear implant. The research platform was designed to be user-friendly, not to require programming skills (for other non-technically trained users), and at the same time to allow for great flexibility. The research platform allows for almost all types of psychophysical and sound-processing research. It is capable of handling electrical stimuli for cochlear implants or acoustical sounds and exploits nearly all the possibilities of the Nucleus CI24 implant. This research platform is used in the experiments of all further chapters. The interface presented to the subjects was also extended with an animated graphical mode to allow for testing children from the age of 2.5 years.

In **Chapter 3** the smallest detectable changes in place pitch are measured for multi-channel stimuli in cochlear implant subjects. The number of channels in the stimuli was varied but no significant difference was found between single or multiple-channel stimuli. These results are consistent with a model for place pitch discrimination which is based upon the edges of the excitation pattern.

In **Chapter 4** the combination of temporal and spatial pitch cues in a pitch ranking task is assessed. This is assessed in normal-hearing subjects using a model that allows for independent control over place and temporal pitch cues. We show that temporal and place pitch cues form a combined percept in a pitch ranking task. For consistent temporal and place pitch information these cues integrate constructively. However, for inconsistent cues, one cue can cancel the other cue.

Chapter 5 investigates the relative contributions of place and temporal pitch cues to F0-discrimination in cochlear implantees. The experiment investigates F0-discrimination using the current sound processing scheme for different filter banks and envelope detection methods. The results indicate that both temporal and place pitch cues contribute to F0-discrimination and their relative power depends on the filter bank used as well as on the stimuli. Secondly, the results also show that the place pitch of a complex multi-channel stimulus corresponds to the centroid of the stimulation pattern along the electrode array. And the results also indicate that the temporal pitch cues are maximal when the modulation depth related to F0 is as deep as possible and in phase across the channels.

In **Chapter 6** the correspondence of pitch perception between cochlear implantees and normal-hearing persons with two different acoustic models is investigated. The goal of the study was to be able to replicate a portion of the data of chapter 5, which was obtained with cochlear implantees, in normal-hearing persons. The results show that sounds processed through noise-band vocoders can be analogous in pitch between normal-hearing subjects and CI subjects. The design of the resynthesis filter bank of the noise-band vocoder however has to match the physiology of the cochlear implant system.

In **Chapter 7**, a new signal processing scheme is designed to optimise pitch perception in cochlear implantees based upon the results found in the previous chapters. This new sound processing scheme is compared to the current clinically most used sound processing scheme with respect to F0-discrimination, melody recognition, and musical interval labelling. The results show that the new sound processing scheme improves F0-discrimination of musical notes, allows for better melody recognition, and allows for a better accuracy in musical interval labelling.

Finally, **Chapter 8** provides the overall conclusions of the thesis as well as some suggestions for further research.

1.5 Conclusion

Severe hearing impairment or deafness often leads to social isolation of the patient and limits the patient's perspectives both professionally as privately. Fortunately, cochlear implants can provide a means for oral communication in most of the deaf. However, the sound that cochlear implant systems provide is still very different from

the sound normal-hearing persons perceive. One of the major issues with current cochlear implants is the poor pitch perception that patients demonstrate. Pitch is, however, important for suprasegmental information in western languages (such as intonation) and also for music perception. Secondly, pitch is important for speech perception in tonal languages, such as Mandarin Chinese, where the tonal language speakers constitute an enormous potential market. Also pitch is used in every day life to separate the sounds we want to hear from the interfering noise. This leads to a demand for better pitch perception in cochlear implant subjects.

After cochlear implants were introduced in the beginning of this introductory chapter, pitch perception in normal-hearing persons was discussed. This is because the mechanisms and performance of normal-hearing persons are taken as the reference to which the cochlear implant results are compared. This thesis also investigates the perception of pitch in normal-hearing subjects together with the analogy in cochlear implants. The third section of this chapter described the poor performance of CI subjects in music and pitch perception. It also described the state-of-the-art on pitch perception mechanisms in CI subjects. It was shown that two pitch mechanisms are used by cochlear implantees: place pitch and temporal pitch. Unfortunately, F0-discrimination performance using these mechanisms is very poor compared to normal-hearing performance.

The final section of this chapter highlighted the major research questions and goals of this thesis and outlined the structure of the following chapters.

Chapter 2

The research platform: APEX

J. Laneau, B. Boets, M. Moonen, A. van Wieringen, and J. Wouters. A flexible auditory research platform using acoustic or electric stimuli for adults and young children. *Journal of Neuroscience Methods,* 142:131-136, March 2005.

Because pitch is *in se* a subjective sensation in man, assessing pitch perception is only possible through experiments with humans. This raises the need for an experimental setup that allows testing the subjects in a controlled way so that the results can be reproduced. This experimental setup should preferably be versatile and user-friendly at the same time. It should also allow the researcher to stimulate the cochlear implant directly to have maximal control over the presented stimuli while exploiting the capabilities of the implant as much as possible. Moreover, because in this thesis the results of the cochlear implant subjects will be compared to the results of normal-hearing subjects on the same tasks, the experimental setup should allow for easy extension of the experiments to acoustical stimulation to normal-hearing subjects.

In the following chapter, a research platform, referred to as APEX (Application for PsychoElectrical eXperiments), is developed that allows the researcher to set up arbitrary auditory psychophysical experiments. The platform takes care of automatic stimulus presentation and collection of the subjects' responses. The setup is capable

of presenting electrical stimuli through cochlear implants or acoustical stimuli to the subjects, either cochlear implantees, hearing impaired, or normal-hearing people. The platform currently supports LAURA, Nucleus CI22 and Nucleus CI24 cochlear implants. The graphical user interface for the subjects has been extended to allow for testing very young children (age down to 2.5 years), by embedding the psychophysical procedures in a computer game. This was done as part of a collaborative study not included in this thesis [Boets et al., 2005]. The research platform is available to other international research groups free of charge.

2.1 Introduction

The introduction of the PC as a regulator for experiments into the area of auditory psychophysics resulted in better controlled experiments and as such to better science. This chapter describes a PC-based research platform for controlling stimulus presentation and collecting subject responses that enables researchers to set up auditory experiments in a user-friendly way. The stimuli can be either acoustical or electrical for cochlear implants, and the graphical user interface for the subjects is adapted for testing both adults and young children.

Software running on digital hardware has enabled researchers to set up more complex test designs, controlling many different factors and varying more parameters than would be feasible for a human controller. Another benefit of software driven experiments is the high precision of stimulus presentation and response collection and the reduction of possible errors during an experiment.

Because of these advantages, psychophysical experiments are currently almost always controlled by software. Most of this software is specifically designed for one particular experiment, but some software packages enable the researcher to set up a large variety of experiments in a user-friendly way. This prevents the researcher from having to code new software for every new experiment. Recently, a number of commercial packages, consisting of software or hardware or both, have become available for performing psychoacoustical experiments.

However, there are still a number of shortcomings to the software that is currently available for auditory experiments. Firstly, most software packages are only capable of presenting acoustic stimuli, while with the increasing number of cochlear implant patients, more and more research is being devoted to electrical auditory perception [Zeng, 2004]. Secondly, most existing packages are designed to test adults and provide the subjects with a relatively dull graphical user interface and limited reinforcement. This excludes testing of young children since their attention fades after a few trials.

This chapter describes a software package called APEX (Application for Psycho Electrical eXperiments) running under the Windows operating system for setting up and performing auditory experiments with acoustical stimuli or electrical stimuli for cochlear implants. The graphical user interface can be set up to a relatively sober mode for testing adults or a highly animated mode for testing children. The software was developed at the Laboratory for Experimental Otorhinolaryngology of the Katholieke Universiteit Leuven, based upon software by Geurts and Wouters [2000]

for the LAURA cochlear implant. It is offered as a free service to the community for research or educational purposes. Information on how to obtain the research platform is given below.

2.2 Overview of the software

After defining the research hypothesis and the general outline of a particular study, the process of designing an auditory experiment can be divided into four steps: 1) stimulus creation, 2) definition of the experimental procedure, 3) running the experiment, and 4) analysis of the collected responses. Each step is supported by APEX software and will be briefly discussed below.

2.2.1 Stimuli

Currently APEX is capable of handling stimuli for four different devices: acoustical stimuli via a PC-soundcard, and electrical stimuli routed through three cochlear implant systems supported by Cochlear Ltd.: Nucleus CI22, Nucleus CI24, and LAURA.

1. Acoustical stimuli consist of standard windows PCM (Pulse Code Modulation) wave files and are presented through the PC-soundcard. The stimuli can contain one or two channels (stereo) with 16 bit resolution and can be sampled at any sampling frequency supported by the soundcard.

2. For electrical stimuli presented through the LAURA implant, a specific file format was designed taking into account the limitations of the implant [Geurts and Wouters, 2000]. The amplitude of the current of LAURA stimuli is always defined relative to the dynamic range of the subject for the different channels and the mapping to the actual current intensity is done in real time by the APEX software. Presentation of electrical stimuli to the LAURA implant requires dedicated hardware mainly consisting of a DSP (TI C30) and a LAURA headset [Geurts and Wouters, 2000].

3. For the Nucleus implant two different file types are supported: NIC1 and NICStream [Cochlear Ltd, 2002]. NIC1 files can contain arbitrary pulse trains (any pulse can be presented at any time) but the file format restricts the total length of the stimulus to approximately 15000 pulses. The NICStream file format supports stimuli of virtually unlimited length but restricts the stimuli to monopolar mode with constant pulse rate and constant pulse shape. The stimuli for the Nucleus implant (of either file format) can be presented using the standard clinical Nucleus fitting hardware and a SPrint speech processor. The amplitude of the current of the pulses in the NICStream format can be stored directly in absolute current intensity or in a relative proportion of the dynamic range. When the amplitude of the stimuli is defined in a relative way, the stimuli can be presented to different subjects by mapping the amplitude of each stimulus between threshold and most-comfortable level of the patient for the particular channels. Mapping of current amplitude is done automatically in APEX by retrieving patient specific data through an SQL-connection to the

clinical fitting software of Cochlear. For the NIC1 file format the amplitude is always defined in absolute current intensity.

All stimuli have to be created offline with APEX and are stored on hard disk. This allows for generating very complex stimuli which may require relatively large computation times. This also enables the researcher to verify all stimuli before running the experiment. Before starting an experiment, APEX checks all stimuli for file-corruption and compatibility of the stimuli with the present hardware. An extensive list of Matlab functions is available for creating, visualizing and storing stimuli for all file formats.

2.2.2 Experimental procedure

APEX was designed to be user-friendly and not to require any programming skills for creating and setting up an experiment. Experimental procedures and parameter values are defined in simple configuration text files that are interpreted by the APEX program. The syntax of the experiment configuration is straightforward and easy to read.

Currently APEX is capable of performing five different types of experiments: 1) constant stimuli experiments, 2) adaptive procedure experiments, 3) closed and 4) open set identification, and 5) adjustment procedure experiments. Both the constant stimuli procedure and the adaptive procedure present the stimuli in an N-interval, M-alternative-forced-choice paradigm where N and M are arbitrary with M being smaller or equal to N. Whereas the constant stimuli procedure presents every test trial a fixed number of times, the adaptive procedure uses the x-up, y-down procedure [Levitt, 1971] to reach a certain asymptotic level of performance. Closed set identification allows the researcher to set up a list of possible answers from which the subject has to select one after each trial. During open set identification the subjects can respond with any possible answer. With the adjustment procedure, the researcher can instruct the subject to adjust a property of the sound to match a predefined percept (e.g. matching the percept of a companion stimulus). During one step of an adjustment procedure the variable test signal is presented once in sequence with a fixed reference signal. The reference signal can optionally be repeated a number of times before and after the test signal. The subject then indicates how to adjust the variable test signal, direction and step size (for instance large step, normal step or small step) and a next trial is presented to the subject with an adjusted test signal. This procedure is repeated until the subject is satisfied with the correspondence between reference and test signal.

2.2.3 Running the experiment

After the stimuli have been created and the experimental procedure has been defined by the researcher, subjects can perform the experiment. The subject sits in front of a graphical user interface as in Figure 2.1 and is presented with the sequence of trials. After each trial the subject enters his response. This can be done using the PC-mouse, keyboard, pen-pad or touch screen. Optional feedback can be provided after each response.

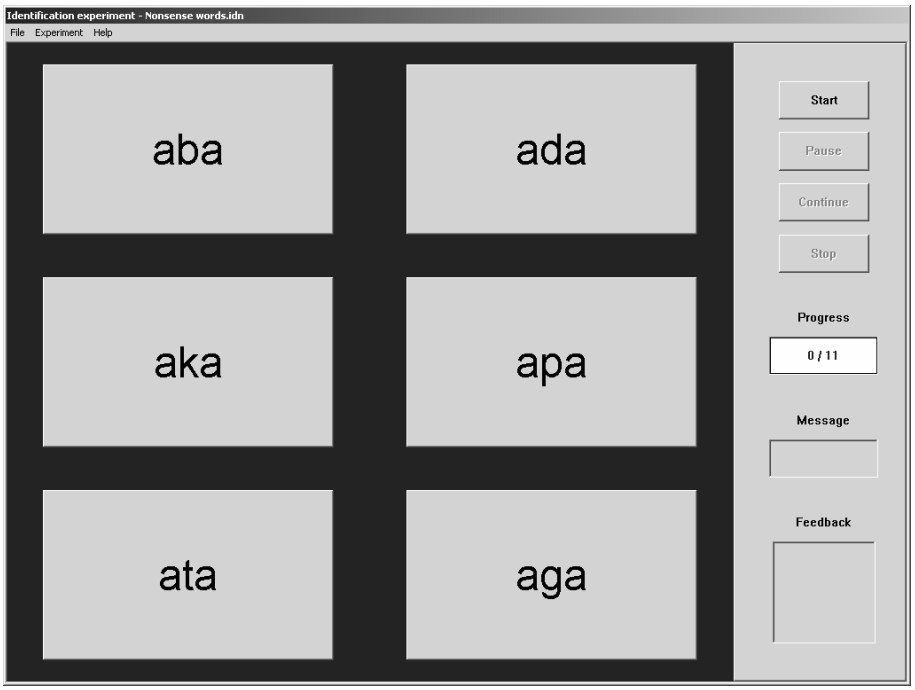


Figure 2.1: Screenshot of an example of a closed set identification experiment using APEX. The subject is presented with a Vowel-Consonant-Vowel stimulus and has to select the correct consonant from the list of six possible answers.

2.2.4 Analysis of results

In order to permanently associate the experimental parameters with the obtained results, the program collects all responses and outputs them to the same text file that has been used for defining the experiment. The resulting output can be the complete sequence of trials consisting of presented stimuli and responses, or it can be a more compact form depending on the task. For instance, for constant stimuli or adaptive procedures the program can output the obtained psychometric function. In addition the mean of the last reversals can be printed for the adaptive procedure, where the number of reversals averaged can be set by the experimenter. For the closed set identification the program computes and prints the confusion matrix. These data can then easily be imported into statistical software packages for further analysis of the results.

2.3 Internal structure

The program was designed in an object oriented way using C++ with MFC (Microsoft Foundation Classes). The object oriented programming allows for a high level of abstraction and easily manageable code. The summary class diagram of the application is shown in Figure 2.2. The main application loads the experiment configuration file and creates a new document object (*ApexDocument*) based upon the configuration file. The *ApexDocument* class is derived from the general

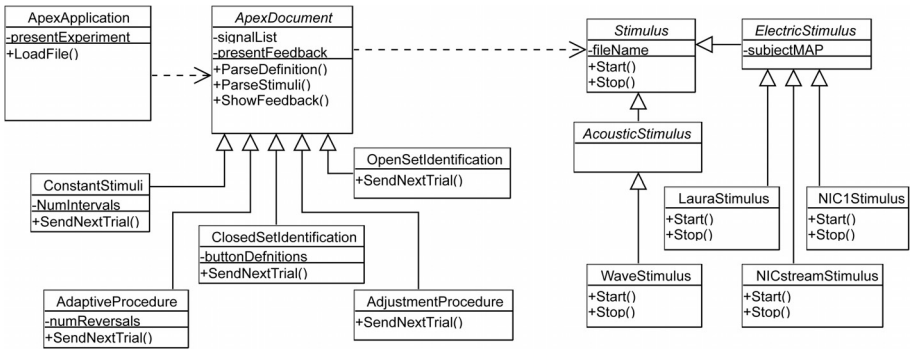


Figure 2.2: UML (Unified Modeling Language) scheme of the internal structure of the APEX source code. See text for details.

CDocument class from MFC and embodies a large number of routines that are common to all experimental procedures, such as displaying start buttons, presenting feedback, parsing the definition of the experiment, checking stimulus files, etc ….

For each particular experimental procedure (adaptive, constant stimuli, …) a specific class is derived from this general *ApexDocument* class. These derived classes generally extend the base class with functions that, for instance, define which stimuli have to be presented in the next trial and in which order, depending on previous responses. The *ApexDocument* class has an extensive list of functions so that implementing new procedures requires minimal coding.

The general ApexDocument class also contains a list of stimuli. These stimuli are treated as abstract stimuli without device or file format specification. For each type of stimulus a specific class is derived from the abstract *Stimulus* class. These classes implement the actual playback of the auditory signals and are device and file format dependent. This structure assures that when adding support for a new kind of stimulus or device, all experimental procedures immediately support this new type of stimulus. And also, when adding a new experimental procedure to the program, the procedure immediately works for all types of stimuli or devices.

2.4 Animated mode

The graphical user interface has been extended to allow testing of very young children. Currently, this animated mode is applicable for constant stimuli experiments, adaptive experiments and closed set identification experiments. In general, children - and especially preschoolers - are much harder to test because of the cognitive demands and the dull repetitive structure of psychophysical tasks. In order to make the relatively tedious psychophysical tests more interesting and child-friendly we integrated them in an interactive video game. This concept of testing was based upon earlier work of Soderquist and Shilling [1992] and Allen and colleagues [Allen et al., 1989; Wightman et al., 1989].

Because the most important condition for successful psychophysical testing of young children is the transformation of the abstract meaningless acoustical signal into a concrete and well known ‘daily life signal’, we have devised three short stories. These stories were intended to give a psychological and naturally fitting

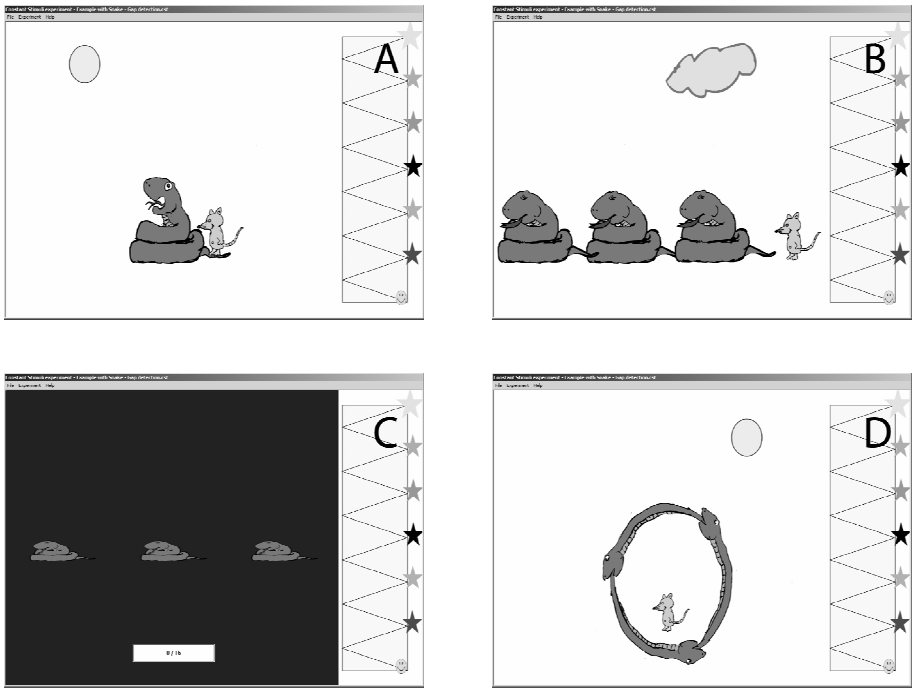


Figure 2.3: Screenshots of the graphical user interface for testing young children. The present screenshots show the scenario using the sleeping snakes and the teasing mouse. Panels A and B are taken from the introductory movie. Panel C shows the graphical user interface for a 3I-3AFC procedure with reinforcement. Panel D shows a screenshot taken from the concluding movie.

meaning to a broad set of possible stimuli. Every story was worked out through a collection of animated cartoons. An introductory movie tells a story and introduces the specific target sound in a game-like manner. During tests with multiple-sound-interval trials, every interval is visually represented on the screen by a funny character (of for instance an animal, or a car). The characters are animated synchronously with the presentation of the corresponding sound interval. The experimenter can have manual control over the initiation of consecutive trials and as such is able to initiate a trial only when the child is attentive and looking at the monitor.

The child's task is to choose, by pointing to the (touch) screen, which of the characters corresponds to the target sound. The child can be given an unlimited time in which to respond. Immediately after the child's response, visual feedback is given in the form of an animated cartoon. A correct response is reinforced with a spectacular movement of the selected character. In case of an incorrect response, feedback (presented through animation of the misidentified character) is much more modest. During the sequence of trials constituting of a single run, APEX also provides a more global reinforcement by adding smiley faces to a rising ladder structure for every correct response. Every incorrect response is optionally punished by removing a smiley from the ladder structure. At the end of the run the number of

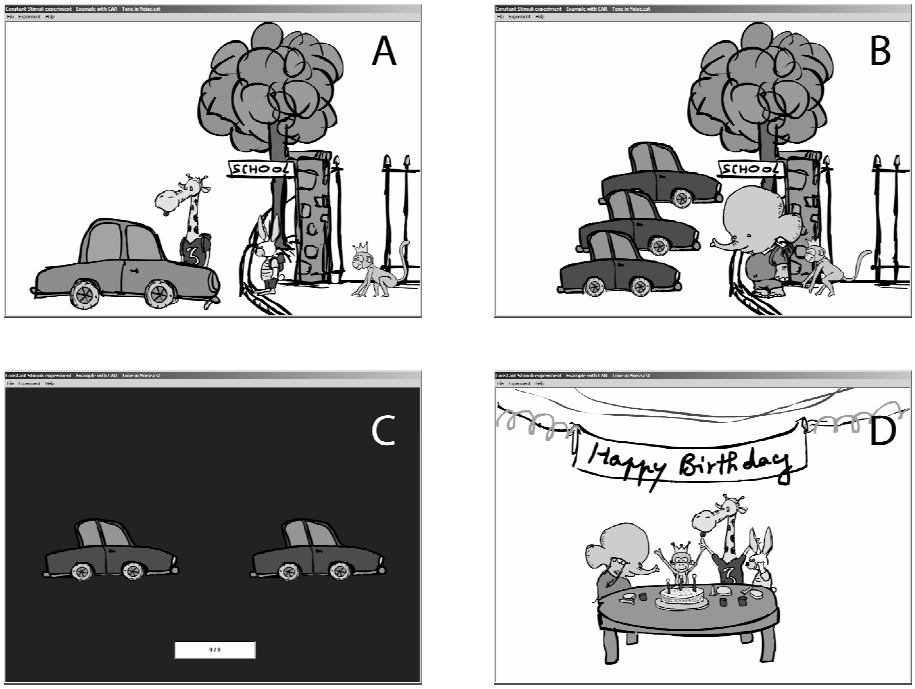


Figure 2.4: Screenshots of the graphical user interface for testing young children. The present screenshots show the scenario using the cars and the monkey. Panels A and B are taken from the introductory movie. Panel C shows the graphical user interface for a 2I-2AFC procedure without reinforcement. Panel D shows a screenshot taken from the concluding movie.

smiley faces may be evaluated and the child could be rewarded accordingly. The ladder structure with the smileys is continuously presented to the subject at the right of the computer screen. To round off the experiment and as a final reward for the child's achievement, an animated closing movie can be shown.

The intro and closing movies are optional and can be skipped for children that require less conditioning. Also the reinforcement can be adjusted by leaving out feedback or omitting the ladder structure with the smileys.

At present, APEX contains three sets of animated cartoons, that have been used for gap-in-noise detection, frequency modulation detection and tone-in-noise detection experiments with normal-hearing children [Boets et al., 2005] and for speech tests (word recognition and phoneme discrimination). Of course, these scenarios are applicable to any kind of experiments like frequency and intensity discrimination, amplitude modulation detection, etc ….. If deemed necessary for appropriate conditioning of the child, it is possible to invent a new introductory story (for the optional intro and closing movies) or adjust the existing ones to the particular task at hand.

In the first scenario, the introductory movie shows a teasing mouse that is trying to wake up three sleeping snakes because she wants to play with them (see Figure 2.3).

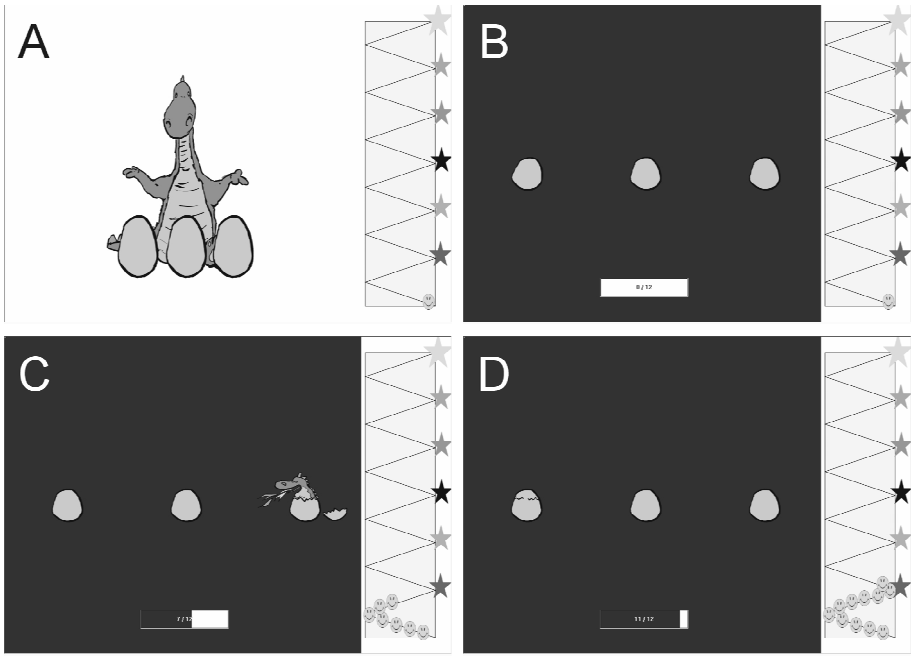


Figure 2.5: Screenshots of the graphical user interface for testing young children. The present screenshots show the scenario using the dragon eggs procedure with three intervals. Panel A depicts a frame of the introductory movie showing the mother dragon trying to figure out which egg is ready to hatch. Panel B depicts the graphical user interface at the start of the test. Each stimulus interval is represented by a dragon egg. At the right hand side of the user interface the ladder structure for reinforcement is shown. Panel C depicts the positive feedback after a correct answer at the seventh trial (indicated in the progress bar at the bottom of the user interface). Note the increase in number of smileys in the ladder structure to reinforce the children. Panel D depicts the negative feedback after an incorrect answer.

The movie explains that it is possible to predict which snake is going to wake up by listening to the specific sound 'ssss-ssss'. During stimulus presentation the snake moves as if it is breathing very deeply. A correct response is reinforced by the snake waking up with a spectacular somersault. An incorrect response results only in a modest movement of the snake's tail. The concluding movie shows all the snakes waking up and playing enthusiastically with the mouse.

The second scenario (see Figure 2.4) starts with an introductory movie about a little monkey. The monkey is having his birthday-party and is waiting desperately at the school gate for his father to pick him up and bring him to his party. Three cars arrive at the school, while the movie explains that it is possible to identify the father's car since he always honks his horn. During stimulus presentation every car is driving (wobbling) synchronously with the corresponding noise sound. A correct response is reinforced with a fragment of the monkey's father jumping enthusiastically out his car. An incorrect response results only in a short appearance of an elephant trunk out

of the roof of the car. In the concluding movie the child can see the little monkey finally arriving at his birthday party.

The third scenario (see Figure 2.5) starts with an introductory movie about a mother dragon sitting in front of three eggs, explaining that it is possible to know which egg is ready to hatch, as you can hear the baby dragon crying inside. During stimulus presentation each egg shakes synchronously with the corresponding sound signal. After a correct response a cute fire spewing baby dragon jumps out of the selected egg. After an incorrect response the selected egg only shows a small crack on its shell. The concluding closing movie shows three little baby dragons breaking out of their eggs and spewing fire.

All animations used in APEX (intro and closing movies, the animated buttons, the feedback and reinforcement) are created using vector-based animations using Macromedia Flash software. This allows the program to easily resize the animations without loss of quality for different screen resolutions or for different screen layouts of the various experimental procedures. Moreover, new animations can easily be created by the researcher and incorporated into APEX.

2.5 Outcomes

The APEX software has been used in numerous experiments in different research groups. The results of these experiments have resulted in more than 15 peer-reviewed international publications [Wouters et al., 1998; Geurts and Wouters, 1999; van Wieringen and Wouters, 1999a; van Wieringen and Wouters, 1999b; Carlyon et al., 2000; Geurts and Wouters, 2001; van Wieringen and Wouters, 2001; Van Ingelghem et al., 2001; Carlyon et al., 2002; van Wieringen et al., 2003; Laneau and Wouters, 2004; Laneau et al., 2004; Geurts and Wouters, 2004; Laneau et al., 2005; van Wieringen et al., 2005].

The accuracy of the test protocols for children using the animations has been validated in normal-hearing children [Boets et al., 2005] and in pilot tests with deaf children implanted with a cochlear implant. These experiments have shown that identification experiments are feasible from 2.5 years of age. The psychophysical discrimination tasks were successfully carried out with 4 year olds.

2.6 Summary

A versatile and easy to use research platform has been developed that allows for running auditory experiments with either acoustical stimulation or electrical stimulation through different cochlear implant devices. The platform allows to test very young children because the integration of the psychophysical procedures within a child-friendly computer game with an extensive reinforcing system contributes significantly to keeping the child motivated and attentive. This platform will be used in all experiments of the following chapters.

Readers interested in obtaining APEX are requested to contact the author or through the APEX website (http://www.kuleuven.ac.be/exporl/Lab/Research/APEX/). APEX is generally available free of charge subject to compliance with the license

agreement. The main requirement of this license agreement is proper citation to the manuscripts describing the research platform.

Chapter 3

Multi-channel place pitch sensitivity in cochlear implant subjects

J. Laneau and J. Wouters. Multi-channel place pitch sensitivity in cochlear implant recipients. *Journal of the Association for Research in Otolaryngology,* 5(3):285-294, September 2004.

Cochlear implant recipients perceive a rise in pitch when the site of stimulation is moved from the apex towards the base. The place pitch sensitivity is typically measured using the stimulation of single channels. However, all current cochlear implant devices stimulate multiple channels simultaneously or with pulses temporally interleaved. The primary goal of the study in this chapter is to test whether the sensitivity of a cochlear implant recipient to place pitch changes improves or deteriorates when the number of active channels is increased, compared to stimulation with only one active channel.

In this chapter an experiment was set up where place pitch sensitivity was recorded in four Nucleus CI24 subjects as a function of the number of active channels (from 1

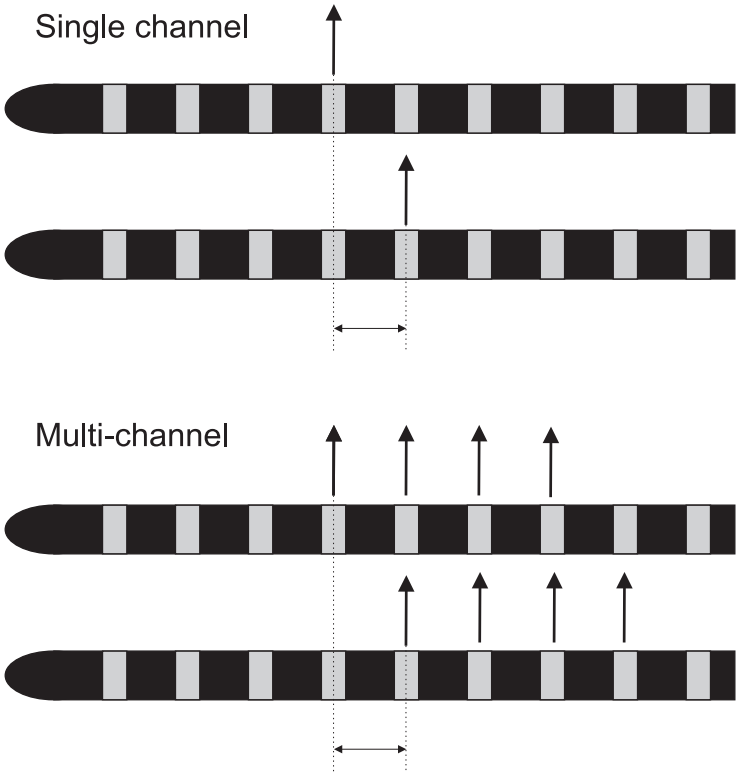


Figure 3.1: Schematic overview of the stimuli used in the experiment. The top panel shows two single channel stimuli. The bottom panel shows two four channel stimuli. The arrows indicate which electrode contacts are activated. The shift in place is the same (one inter-electrode spacing) between the two single channel stimuli as between the two four channel stimuli.

to 8). Just noticeable differences were estimated from a constant stimuli 2AFC pitch ranking experiment with roving loudness. Reference and comparison stimuli contained the same number of active channels but were shifted one or two electrodes towards the base or towards the apex. The place pitch sensitivity was measured using monopolar stimulation at two locations along the electrode array. To minimize cues related to loudness, the multi-channel stimuli were loudness balanced relative to the single-channel stimuli presented at comfortable level.

The results of the experiment show that the number of active channels did *not* affect place pitch sensitivity. This is consistent with models that compare the edges of the excitation pattern or the centroids of the excitation pattern, both irrespective of the overlap between excitation patterns. There was a significant difference in sensitivity to place pitch among subjects. The average just noticeable differences of place pitch, extrapolated from a fitting procedure, for the subjects ranged from 0.25 mm to 0.46 mm.

3.1 Introduction

Studies of place pitch sensitivity have typically revealed a large inter-subject variability ranging from near perfect discrimination of adjacent channels up to weak discrimination of channels at opposite extremes of the electrode array [Nelson et al., 1995]. This large range of JNDs and inter-subject variability has been corroborated by other studies [Busby and Clark, 1996; Collins et al., 1997; Zwolan et al., 1997; McKay et al., 1999; Pfingst et al., 1999]. Reducing the loudness of the stimuli tends to decrease place pitch sensitivity or electrode discrimination, though the effect is small [McKay et al., 1999] and might not be present in all subjects [Pfingst et al., 1999]. Also, random roving of the loudness decreases place pitch sensitivity [Henry et al., 2000].

The perceived magnitude of place pitch has been proven to be related to the center of gravity of the forward masking pattern for single-channel stimuli, and thus related to the center of gravity of the excitation pattern [Cohen et al., 1996a]. Furthermore, intermediate place pitches can be elicited by simultaneous stimulation of adjacent channels [Townshend et al., 1987], or non-simultaneous stimulation with pulses temporally interleaved [McDermott and McKay, 1994].

Although place pitch has not been studied extensively in a musical context, McKay et al. studied musical interval identification using single-channel pulse trains in one musically trained subject [McDermott and McKay, 1997]. They found that musical interval identification is possible using place pitch although a large variability existed in the responses, and that the size of the interval is approximately linearly related to distance along the electrode array.

Most studies that report on place pitch discrimination use bipolar stimulation, where both the active and the return electrode are intra-cochlear. However, in current clinical practice more and more cochlear implant devices are stimulated in monopolar mode where the return electrode is distant and located outside the cochlea [Pfingst et al., 2001]. Monopolar stimulation requires lower stimulation levels, and can lead to better speech perception in cochlear implant recipients [Zwolan et al., 1996]. In the present study place pitch discrimination was determined using monopolar stimulation.

Studies of pitch sensitivity in cochlear implant subjects, based on either temporal or place pitch cues, typically involve simple stimuli, usually only stimulating one channel or electrode at a time. However, current speech processors stimulate a number or even all channels simultaneously or with pulses temporally interleaved depending on the stimulation strategy used. Few studies have measured the effect of multiple-channel activation on pitch sensitivity. In a study by Geurts and Wouters [2001] the discrimination of the fundamental frequency of synthetic vowels was measured after processing with a CIS strategy. They also found that both temporal and place pitch cues can contribute to the discrimination of the fundamental frequency of synthetic vowels. In an additional temporal pitch discrimination experiment they also concluded that multiple-channel stimulation improves the temporal pitch sensitivity. However, in their experiment the loudness of the multi-channel stimuli was higher than the loudness of the single-channel stimuli. This might have contributed to a better discrimination for the multi-channel stimuli

because rate discrimination improves as loudness increases [Morris and Pfingst, 2000].

In this chapter we measure the effect of multiple-channel stimulation on place pitch sensitivity (see Figure 3.1). We hypothesize that place pitch sensitivity will be affected by the number of active channels. Place pitch sensitivity might decrease because of the increased overlap in excitation patterns. Similarly, place pitch sensitivity might decrease with increasing number of active channels if the sensitivity is based upon the peakedness of the excitation pattern. In contrast, the distance between the channels that are not common to both stimuli in one trial becomes larger as the number of active channels increases, although the overall shift of active channels is fixed. And then this larger distance might improve place pitch sensitivity. If place pitch discrimination is based upon size of the difference in excitation at the edges or flanks of the excitation patterns, it will be unaffected by the number of active channels.

3.2 Methods

A pitch ranking experiment using single and multi-channel stimuli examined whether the number of active channels affects the subjects' sensitivity to place pitch. The channels were stimulated in monopolar mode. Just noticeable differences (JND) were estimated from this pitch ranking experiment and expressed as a function of the number of active channels. The multi-channel stimuli were loudness balanced to the single-channel stimuli in order to minimize cues related to loudness. The subjects were trained with the psychophysical task prior to the start of the experiment in order to minimize training effects. Apex (see chapter 2) was used to run all experiments.

3.2.1 Subjects

Table 3.1: Relevant information about each of the subjects who participated in the experiments. The last column shows the results for a phoneme recognition task with monosyllabic consonant–vowel–consonant (CVC) words of the NVA list [Wouters et al., 1994].

Subject	Duration of profound deafness (y)	Etiology	Implant Type	Clinical speech processing strategy	Age (y)	Implant experi-ence (y;m)	Phoneme recognition (%)
CI1	11	Hereditary	Nucleus CI24R(ST)	Speak	19	1;11	74%
CI2	3	Unknown	Nucleus CI24R(ST)	Speak	21	3;2	86%
CI3	30	Progressive	Nucleus CI24R(CS)	ACE	54	0;8	84%
CI4	32	Progressive	Nucleus CI24R(CS)	ACE	49	0;6	52%

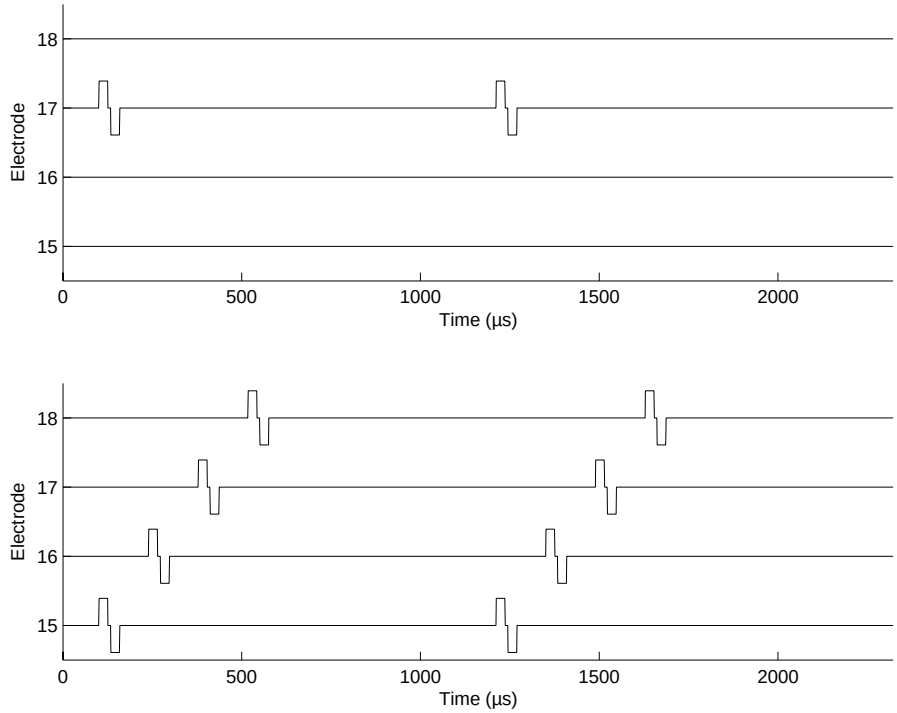


Figure 3.2: Amplitude-time waveforms of parts of two different stimuli, both centered on electrode 17. Electrodes are numbered from base to apex. The upper panel displays a part of a single-channel stimulus. The lower panel displays a part of a four-channel stimulus.

Four post-lingually deafened users of the Nucleus CI24 cochlear implant participated in this study. All subjects performed all tests and were paid for their collaboration. Some relevant details about the subjects can be found in Table 3.1. Subjects CI1 and CI2 are implanted with the straight version of the Nucleus electrode array CI24R(ST) Nucleus24k. Subjects CI3 and CI4 are implanted with the peri-modiolar version of the Nucleus electrode array CI24R(CS) Nucleus24 Contour. All subjects achieve relatively good speech recognition.

3.2.2 Stimuli

The stimuli in the experiments consisted of constant amplitude biphasic pulse trains with phase duration and inter-phase gap set to 25 µs and 8 µs respectively. The pulse rate was 900 pulses per second (pps) per channel. The stimulation mode was monopolar with both return electrodes active (MP 1+2), so every stimulation channel consisted of one intra-cochlear electrode and two extra-cochlear electrodes. All stimuli were 500 ms in duration with a 500 ms silent gap in between the two stimuli of a test trial.

The single-channel stimuli consisted of 450 pulses of equal amplitude presented to the desired channel. The multi-channel stimuli consisted of several adjacent single-channel stimuli presented with pulses temporally interleaved. This was done by

delaying every single-channel pulse train added to the stimulus by 138 μs (equivalent to the inter pulse interval of an ACE strategy with 8 maxima at 900 pps per channel) relative to the last added pulse train. This allows a maximum of 8 simultaneously active pulse trains. Multi-channel stimuli are indicated by their center electrode for stimuli containing an odd number of active channels. For stimuli containing an even number of channels the stimuli are indicated by the more apical electrode of the middle two electrodes. The stimuli used in the present study were centered on electrodes 13 or 17 and the number of active intra-cochlear channels was 1, 2, 3, 4, 5, or 8 depending on the condition. The electrodes are numbered from base to apex, according to clinical practice of Cochlear devices. Figure 3.2 displays a detail of two examples of stimuli: a single-channel stimulus and a 4-channel stimulus both centered on electrode 17.

The amplitude of the single-channel stimuli was at C-level (comfortable loudness level) or 100% of the dynamic range for the respective channel and subject. The amplitude of the multi-channel stimuli was set to a fixed percentage between 0% and 100% of the dynamic range expressed in clinical current units. The clinical current units follow a logarithmic scale and one clinical current unit corresponds to approximately 0.17 dB. The amplitudes of all channels of one multi-channel stimulus were set to the same percentage. This percentage was chosen to make the single and multi-channel stimuli equally loud and was obtained through the loudness balancing procedure discussed further. The amplitude of the pulses to each channel in a multi-channel stimulus was set to the current level that corresponded to this percentage of the dynamic range for the respective channel. Consequently, the amplitude of pulses was constant within a channel but not across channels. Setting the amplitude for each channel at the same percentage of the dynamic range, makes all channels approximately equally loud when presented separately - at least when assuming that the loudness growth functions as a function of dynamic range are similar over neighboring channels. This assumption is also used in the clinical implementation and fitting of speech processors.

The dynamic ranges for each channel and each subject were obtained from the clinically determined hearing threshold (T-level) and C-levels for single channel pulse trains at 900 pps.

To minimize cues related to loudness, random current intensity roving was applied by multiplying each stimulus' amplitude with a random value between 0.85 and 1.1, where the amplitude was expressed in clinical current units relative to the T-level. The roving factor was the same for all channels of a multi-channel stimulus.

3.2.3 Loudness balancing

The multi-channel stimuli were loudness balanced for each subject to the loudness of a single-channel stimulus. This single-channel stimulus was presented at full dynamic range or C-level and at the center of the respective multi-channel stimulus. The loudness balancing procedure was performed for stimuli centered on electrodes 17, 13 and 6 for subjects CI2 and CI4. For subjects CI1 and CI3 the loudness balancing was done for electrodes 17 and 13.

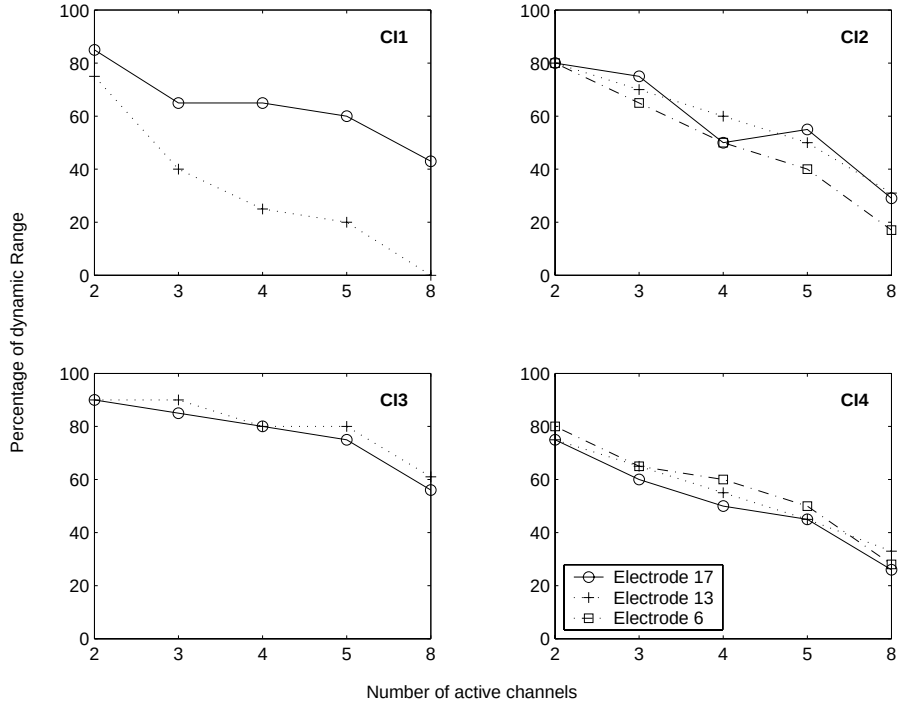


Figure 3.3: Results of the loudness balancing of multi-channel stimuli relative to a single-channel stimulus presented at C-level. Loudness balanced stimuli were always centered on the same electrode. Each panel corresponds to one subject and the different lines within each panel represent results for different center electrodes. The loudness of the multi-channel stimuli is adjusted by setting the amplitude of each channel to a percentage of the dynamic range. The latter percentage is fixed over the different channels that make up the multi-channel stimulus.

The loudness balancing procedure was based on a 2 interval – 2 alternative forced choice (2I2AFC) task. The subject perceived in random order the single-channel stimulus presented at C-level and a multi-channel stimulus of which the amplitude was varied, and was asked to indicate the loudest signal. The amplitude of the multi-channel stimulus was adaptively varied using a 1-up 1-down procedure [Levitt, 1971]. The amplitude of the pulses in the multi-channel stimuli was varied in steps of 5% of the dynamic range for amplitudes between 10 and 100% dynamic range. For dynamic ranges between 0 and 10% the amplitude varied in steps of 1%.

The subjects performed two runs of the adaptive procedure for each center electrode and each number of active channels. At the start of the first run the multi-channel stimulus (presented at 10-20% of the dynamic range) sounded softer than the single-channel stimulus. In contrast, the second run started with the multi-channel stimulus (presented at 80-100% of the dynamic range) sounding louder than the single-channel stimulus. Each run was stopped after 8 reversals. The mean of the 4 last reversals of both runs (and rounded to the accuracy of one of the steps of the

dynamic range used in the loudness balancing) was taken as the amplitude of the multi-channel stimulus after loudness balancing. The average absolute difference between both runs was 0.04%.

The resulting loudness balanced amplitudes, presented as a percentage of the dynamic range, are depicted in Figure 3.3 for each subject and for each center electrode location. The amount of current reduction increased monotonically as a function of the number of active channels for each condition in every subject. For subjects CI2, CI3, and CI4 there was virtually no variation over the different electrode locations. For subject CI1 the multi-channel stimuli centered on electrode 13 required more current reduction for loudness balancing than the multi-channel stimuli centered on electrode 17. The average standard errors over the different electrode locations are 0.17%, 0.04%, 0.01%, and 0.02% of the dynamic range for each subject respectively.

3.2.4 Psychophysical procedure

Place pitch sensitivity was measured at two reference locations along the electrode array (electrodes 17 and 13) and for six values for the numbers of active channels (1, 2, 3, 4, 5 and 8 simultaneously active channels). For subject CI2 a third reference location along the electrode array, electrode 6, was also measured.

Place pitch sensitivity was measured in a pitch ranking experiment using a 2 interval – 2 alternative forced choice task. In one trial, the subjects were presented two stimuli in random order and were instructed to indicate the highest one in pitch and to ignore the loudness of the stimuli. Both stimuli contained the same number of active channels, but one stimulus was centered on the reference electrode and the other stimulus was centered on an electrode located 1 or 2 electrodes more towards the base or 1 or 2 electrodes towards the apex.

The trials were presented in blocks. Each combination of reference electrode location and number of active channels was measured in a separate block of trials, leading to 12 (= 2 electrode locations x 6 numbers of active channels) different blocks per subject. Within each block, the four different shifts (-2, -1, +1, +2) of active channels were repeated ten times, leading to 40 trials per block. Each block of trials was presented twice to the subjects. Accordingly, each point on the psychometric function was based on the mean of 20 trials, and each psychometric function was based on a total of 80 trials per subject.

The amplitude of the pulse trains of all stimuli in one block was adjusted to match the result of the loudness balancing procedure for the reference stimulus. Because center electrode location affected the loudness balancing results only mildly (except for subject CI1) and because the shifts in center electrode location are small within one block, the same loudness adjustment in percentages is used for the comparison stimuli as for the reference stimuli. In addition, any possible remaining loudness effects are expected to be eliminated by the amplitude roving.

3.2.5 Training

Because none of the subjects had prior experience in performing psychophysical tasks, they were trained on three different pitch discrimination tasks prior to the experiment. The subjects were trained to a 2I2AFC pitch ranking experiment. The first two training experiments were electrode ranking tasks where the subjects were presented two stimuli consisting of single-channel pulse trains on different electrodes and had to indicate the highest one. In the first training task, the active electrodes of the two stimuli were 4, 8 or 12 electrodes apart and the amplitude was roved as described before. Feedback was presented immediately after each trial only during this first training task. This test was repeated until the subjects performed the task with a success rate of 100% over 20 consecutive trials. This lasted not longer than 15 minutes.

The second training task was similar to the first training task, but electrodes were spaced more closely together and no feedback was given to the subjects. Feedback was removed from this and further experiments in order to make the subjects use their intuitive pitch sensation rather than any trained pattern they might learn during training with feedback. This test consisted of a 2I-2AFC pitch ranking task and amplitude was roved as described before. The active electrodes were spaced 1, 2 or 3 electrodes apart and the electrode discrimination or place pitch sensitivity was determined at four sites along the electrode array (electrodes 21, 17, 13 and 6).

In the third task, the subjects performed a pitch ranking experiment of complex multi-channel stimuli. This task involved over more than 1000 trials per subject in at least 6 sessions of 1.5h. So plateau performance was expected on pitch ranking tests, although this was not proven explicitly, before the data reported here were gathered.

3.2.6 Data analysis

The data of each block of trials were converted into one single parameter that is a measure of the subjects place pitch sensitivity in that particular condition. A typical result of a block of trials for one subject is depicted in Figure 3.4, where the proportion of times that the reference stimulus was perceived to be higher in pitch than the comparison stimulus is plotted as a function of the shift in location of the center electrode of the comparison stimulus. This proportion was fitted to a normal cumulative distribution function using a Gauss-Newton nonlinear least-squares fitting method. The slope of this fitted normal cumulative distribution function was taken as a performance measure of place pitch sensitivity. This fitting procedure is similar to a linear regression with the data points converted into *d'* values. Just noticeable differences (JND) were derived from the fitted curves as the minimal distance expressed in number of electrodes to obtain 75% correct pitch rankings.

The fitting procedure did not converge for blocks of trials that only contained 100% correctly pitch ranked trials for all shifts in location. For these blocks a 5% error (equivalent to half a trial error because each trial was repeated 10 times) in pitch ranking performance was introduced in the proportion of correctly ranked trials of the smallest shift in location before the fitting procedure, as suggested by MacMillan and Creelman [1991].

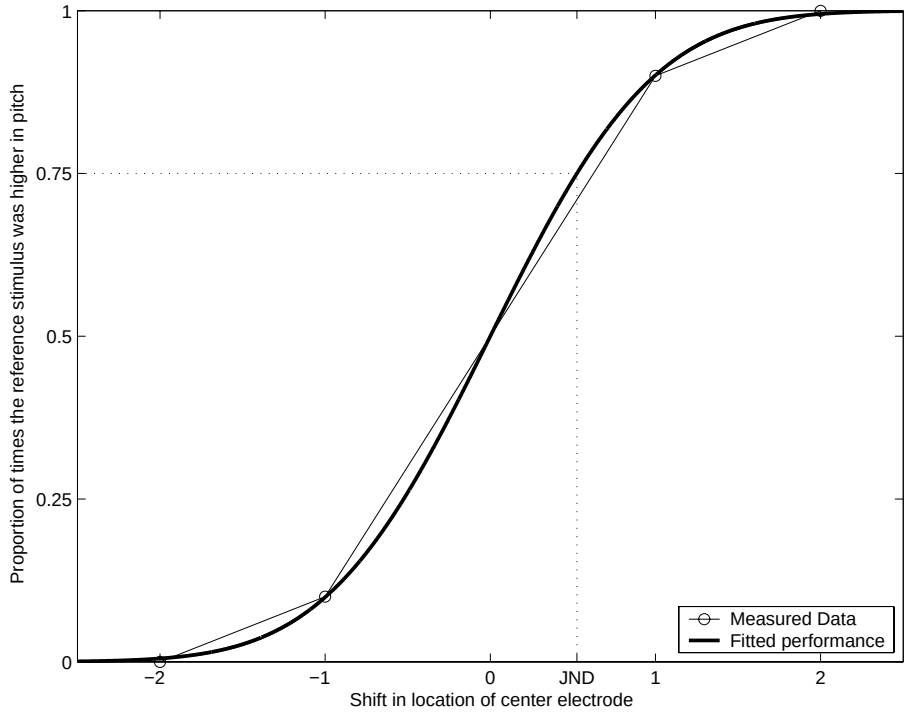


Figure 3.4: Typical result of one block of trials of place pitch sensitivity for one subject. The thin line depicts the proportion of times that the reference stimulus was perceived to be higher in pitch than the comparison stimulus. This proportion is expressed as a function of shift in location along the electrode array of the comparison stimulus relative to the reference stimulus. Negative shifts are displacements towards the apex. The thick line represents the fitted cumulative normal distribution function. The 75% point or JND is indicated on the fitted curve by the dotted line.

3.3 Results

Average JND estimates of the four subjects are depicted in Figure 3.5 as a function of the number of active channels. JNDs are averaged over the two center electrode locations (17 and 13) and both test runs. Subjects scored very well in the discrimination task. The subjects only made between zero and seven errors in pitch ranking per block of 40 trials with an average of 2.1 errors per block, or leading to estimated JNDs that were often smaller than the smallest difference presented to the subjects. Subject CI2 was able to rank all stimuli perfectly in pitch on all trials for all numbers of active channels and for all stimuli centered on electrodes 17 and 13. The JND for subject CI2 was estimated based on the introduction of 5 % errors as described above.

The data were analyzed using a three-factorial repeated-measures analysis of variance (ANOVA). The three factors were center electrode location (electrodes 17 or 13), number of active channels (1, 2, 3, 4, 5, or 8 active channels) and the number

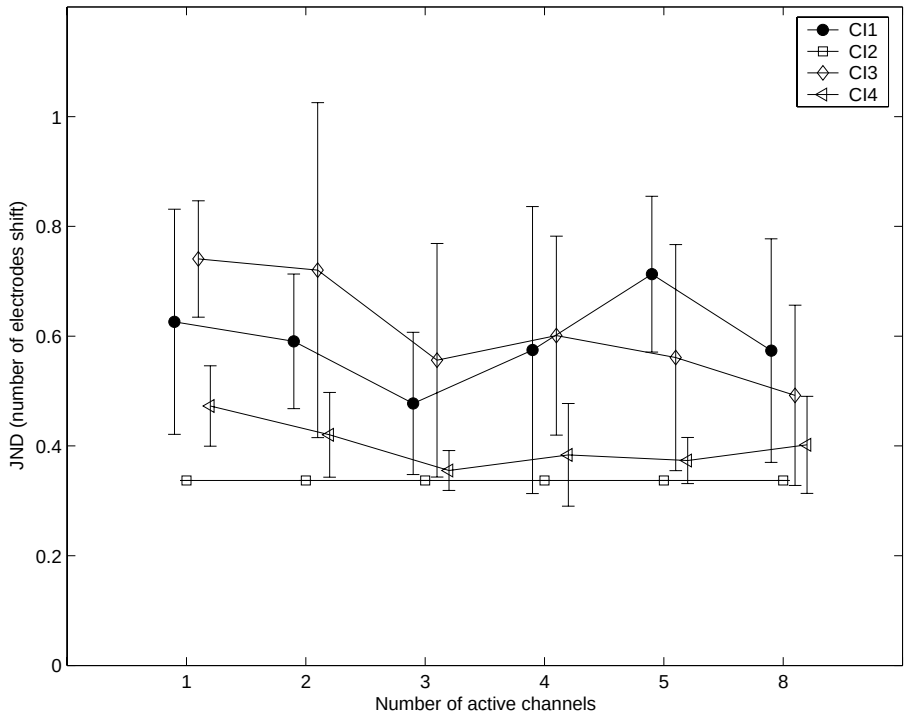


Figure 3.5: The average JND of place pitch sensitivity (based on the fitting procedure) as a function of the number of active channels, for the four subjects. The JND are averaged over the two electrode locations and the two test runs, and they are expressed in number of inter-electrode distances. Error bars indicate ±1 standard deviation. Increasing the number of active channels did not affect the place pitch sensitivity. Subject CI2 had perfect discrimination over all conditions and all runs. The estimated JNDs for CI2 are based upon a 5% error for the smallest tested shift (see text).

of the test run (the first or second test run). No significant effects were found. The effects of the center electrode location ($F(1) = 0.027$; $p = 0.879$) and the effect of test run ($F(1) = 0.113$; $p = 0.758$) were not significant. The effect of the number of active channels was not significant ($F(1) = 2.038$; $p = 0.249$) using the lower-bound statistics. There were no significant interaction effects. There were significant differences in performance between subjects ($F(1) = 50.008$; $p = 0.006$).

The average JND for each subject respectively is 0.59, 0.34, 0.61, 0.40 expressed in electrodes. As electrodes in the Nucleus electrode array are spaced 0.75 mm apart, these JNDs can also be expressed as 2.26 d'/mm, 3.92 d'/mm, 2.19 d'/mm, and 3.33 d'/mm.

Because the data from the training experiments indicated a slightly worse place pitch sensitivity at electrode 6 for subject CI2, this additional condition was presented to subject CI2. On average subject CI2 made 1.5 errors per block, resulting in an average JND of subject CI2 in the extra condition of 0.45 electrodes or 2.96 d'/mm.

The JNDs of the subjects are derived from a fitting procedure as described in the methods section. Most JNDs are interpolated values because they are smaller than the smallest difference actually presented to the subjects.

3.4 Discussion

The obtained channel discrimination JNDs are within the range of JNDs reported for Nucleus CI22 subjects stimulated in BP or BP+1 mode. In the study of Nelson et al. [1995] the average performance of 14 subjects for a 2I-2AFC pitch ranking experiment with loudness balanced stimuli ranged from very poor (0.12 d'/mm) up to excellent (3.16 d'/mm) which translates to JNDs of 11 electrodes and 0.4 electrodes, respectively. Compared to these results, the subjects of the present study all have very good place pitch sensitivity.

Subjects CI3 and CI4 of the present study were implanted with a pre-curved electrode array, the CI24R(CS) Nucleus24 Contour device. JNDs in bipolar mode (BP+1) in three subjects implanted with a prototype pre-curved electrode array ranged between 0.5 and 3 electrodes, measured with a 4I-4AFC discrimination task and with loudness balanced stimuli using an adaptive procedure [Cohen et al., 2001]. The JNDs of subjects CI3 and CI4 of the present study are equivalent to the JNDs of the better performing subject in the study of Cohen et al. The amplitude roving of the stimuli in the present study might have reduced place pitch sensitivity. The JNDs of the subjects in the study of Cohen et al. increased to between 0.7 electrodes and 6 electrodes when loudness was roved between 35 and 80% of the dynamic range [Cohen et al., 2001]. However, the loudness roving in that study was deliberately increased until the task was significantly more difficult for each subject. The amount of loudness roving was smaller in the present study and, consequently, the decrease in place pitch sensitivity due to loudness roving might have been smaller.

The present study measured place pitch sensitivity in monopolar stimulation mode, whereas most studies, including the studies of Nelson et al. [1995] and Cohen et al. [2001], all report channel discrimination in bipolar mode. Few studies have reported channel discrimination in monopolar mode. Townshend et al. [1987] examined channel discrimination in three subjects with monopolar stimulation using a 2I-2AFC procedure. For one subject the pitch varied non-monotonically with distance from the base. The two remaining subjects were able to pitch rank the channels in an orderly way from apex to base. The d' per electrode in those two subjects ranged from 0.1 to 0.9. This corresponds to JNDs ranging from 33 to 3.5 Nucleus electrodes (or ranging from 0.04 d'/mm up to 0.36 d'/mm), respectively, taking into account that the electrodes in the study of Townshend et al. were separated by 2.5 mm. The JNDs reported by Townshend et al. are appreciably higher than the JNDs reported in this study.

Although modeling results and physiologic studies in animals showed that stimulation in monopolar mode excited very broad regions [Black and Clark, 1980], recent data by Nelson et al. [2003a] show that forward masking patterns obtained using monopolar mode can be relatively narrow with relatively steep slopes. They also found that the range of widths of the forward masking patterns is similar for bipolar and monopolar stimulation modes, although the comparison was not done

within the same subjects. The steep slopes of the excitation patterns even when using monopolar stimulation, might explain the excellent place pitch discrimination found in the present study because narrower excitation patterns are assumed to be more easily discriminated.

Although the multi-channel stimuli excite wider and more overlapping populations compared to single-channel stimuli, the discrimination was not diminished. This is consistent with the idea (although it does not conclusively prove) that discrimination is based upon the edges of the excitation pattern, and not upon the overlapping region. This last hypothesis is consistent with the model of Zwicker for frequency discrimination with acoustic stimulation [Zwicker, 1970]. According to Zwicker's model, frequency discrimination depends on the detection of differences in the time-averaged patterns of excitation. Zwicker assumed that a change between stimuli can be detected whenever the excitation patterns differ at any point by more than a criterion value. The size of the JND is predicted to be inversely proportional to the slope of the excitation pattern at its steepest point, and directly proportional to the criterion value. The model of Zwicker is however not capable of explaining data with roving intensity. Lyzenga and Horst [1995] modified the model of Zwicker to account for frequency discrimination using randomly varying intensity by equalizing the average overall level of the excitation patterns of the comparison stimuli prior to the comparison of the two excitation patterns. This model is capable of explaining frequency discrimination when temporal cues are not present in the excitation pattern in the auditory system [Lyzenga and Horst, 1997], as is the case with the stimuli in our experiment. In a study examining the effect of current intensity on electrode discrimination, McKay and colleagues also found that their data were consistent with such a model [McKay et al., 1999].

The results of this study are also consistent with a model that compares the centroids of the excitation pattern elicited by the two stimuli [Cohen et al., 1996a]. Unfortunately, the conditions assessed in this chapter do not allow for testing which of the two models (Zwicker's model or the centroid model) is the most accurate. Future studies, using stimuli where the edge of the stimulation pattern and the centroid of the stimulation pattern are varied independently are necessary to discern between the two models. In the present study the difference in edge of the stimulation patterns or the difference in centroid of the stimulation patterns of the stimuli that were compared within a trial were correlated.

3.5 Conclusions

Place pitch discrimination or site of excitation discrimination was measured in four Nucleus CI24 cochlear implant recipients using single and multi-channel stimuli, whereas previous studies measured single-channel discrimination. The results from this chapter can be summarized as follows:

1. Place pitch sensitivity or electrode discrimination performance for single and multi-channel stimuli in monopolar stimulation mode can be very good. JNDs are smaller than the distance between two adjacent electrodes (which corresponds to 0.75 mm for the Nucleus CI24 device). This performance agrees

with the data of the better performers in other studies for single-channel stimulation in bipolar mode.

2. Place pitch sensitivity is not affected as the number of active channels, or equivalently the width of the excitation pattern, is increased from 1 to 8, even when the excitation patterns of the different stimuli largely overlap.

3. Discrimination of single and multi-channel stimuli based upon place pitch in a cochlear implant is consistent with Zwicker's model [1970] and with a model comparing the centroids of the excitation patterns [Cohen et al., 1996a].

Chapter 4

The combination of consistent and inconsistent place and temporal pitch cues in a pitch ranking task

Most pitch studies have investigated the different pitch mechanisms separately and focus only on one pitch mechanism. In those studies, the other (non-investigated) mechanisms were either ignored, controlled for (by for instance randomization), or cancelled by keeping specific stimulus properties constant. Few studies have actually assessed the combination of the different mechanisms. In this study, the cooperation or combination of two pitch mechanisms is examined: the purely temporal pitch mechanism and the spectral pitch mechanism in NH listeners. The design of the stimuli allows extrapolating the obtained results to the integration of place pitch and temporal pitch in CI subjects.

In this experiment NH subjects were asked to indicate the highest one of two stimuli differing in spectral, purely temporal or both spectral and purely temporal pitch cues. The stimuli consisted of acoustic simulations of the sound cochlear implant subjects hear (see also Chapter 6) and therefore results obtained from the NH subjects may be extrapolated to pitch perception by CI subjects. More specifically the present results may indicate how temporal pitch and place pitch in CI subjects

are integrated, because in the acoustic simulations temporal pitch in CI is matched onto purely temporal pitch in NH and place pitch in CI is mapped onto spectral pitch in NH.

The spectral and purely temporal pitch cues of the stimuli were varied independently. The stimuli the subjects compared in a specific trial differed in purely temporal, spectral or both temporal and place pitch cues combined. The combination of purely temporal and spectral pitch cues could be *consistent* (either both cues indicate a rise in pitch, or both cues indicate a fall in pitch) or *inconsistent* (one cue indicates a rise and the other cue indicates a fall in pitch).

The results show that there is a significant difference between the consistent cue combination and the inconsistent cue combination. This means that the sign of the cues is important and can not be discarded. The results also show constructive integration of consistent cues and destructive interference for inconsistent cues.

In the discussion of this chapter it is shown that the results are consistent with a model that combines both purely temporal and spectral pitch cues optimally into a single percept and that the two cues contribute independently. This may be extrapolated to electrical hearing as indicating that temporal pitch and place pitch in CI subjects are also integrated into a single percept.

4.1 Introduction

In electrical hearing, stimuli differing in place of excitation can be ranked from low to high, when the site of excitation is moved from the apex toward the base of the cochlea. Also, stimuli differing in stimulation rate can be ranked from low to high with increasing stimulation rate. These pitch mechanisms in cochlear implant subjects are called place pitch and temporal pitch, respectively. Tong et al. [Tong et al., 1983], using a multi-dimensional scaling technique, showed that place pitch and temporal pitch are different percepts. In fact, both percepts appeared to lay on an orthogonal grid, or in other words place pitch and temporal pitch are independent mechanisms with separate, independent associated sensations. This finding was later confirmed using an ‘odd-one-out’-detection task where the stimuli differed in both place and temporal pitch [McKay et al., 2000]. In this study both place and temporal cues contributed independently to the ability to discriminate the ‘odd-one-out’ in a set of three stimuli. No interaction was found between the temporal pitch mechanism and the place pitch mechanism.

This finding in CI subjects is supported by our knowledge about the pitch mechanisms in normal-hearing subjects. In normal-hearing subjects, spectral pitch and periodicity pitch also lead to separate and independent sensations [Plomp and Steeneken, 1971].

Furthermore, there is some evidence that CI subjects combine the place pitch sensation and the temporal pitch sensation into a single pitch percept. This occurs for instance when the subjects are asked to rate quantitatively the height of stimuli varying in both place and rate of stimulation [McDermott and McKay, 1997; Fearn and Wolfe, 2000; Zeng, 2002], or when the CI subjects are asked to rank two stimuli in pitch that differ in both place and temporal pitch (see Chapter 5). In the study of

Fearn et al. [2000] it was shown that on average a shift of 2 mm of the place of stimulation corresponds to a doubling of stimulation frequency. However, CI subjects strongly differ among each other in the way they combine both sensations. Where some subjects rely more on temporal pitch cues; other CI subjects base their pitch estimation more on place pitch cues [Fearn and Wolfe, 2000].

The combination of inconsistent pitch cues appears to be particularly complex. Inconsistent pitch cues originate when for instance the place pitch mechanism indicates a rise in pitch but the temporal pitch mechanism indicates a fall in pitch, or vice versa. For normal-hearing subjects, it has been shown that listeners differ in the perceived pitch direction (rising or falling) when presented with the combination of inconsistently combined purely temporal pitch cues and spectral pitch cues [Smoorenburg, 1970; Renken et al., 2004]. Part of the subjects will denote the pitch difference as rising and the other part of the subjects as falling. It is however, unclear what causes this inter-subject difference. In this chapter, we will investigate the combination of spectral and purely temporal pitch cues, especially for consistent and inconsistent cue combination. In order to force the subjects to combine both pitch sensations, a pitch ranking task is used.

The main goal of this chapter is to assess how place pitch and temporal pitch in CI are combined as this information will be used in the following chapters. However, instead of investigating pitch perception in CI subjects directly, pitch is investigated using an acoustic model of CI sensation presented to NH listeners. In chapter 6 this model for pitch perception in CI subjects is further elaborated. In this acoustical model purely temporal pitch percepts and spectral pitch percepts can be elicited independently. Firstly, purely temporal pitch is elicited in normal-hearing subjects when presented with amplitude modulated noise [Burns and Viemeister, 1976] and this relates to the temporal pitch in CI subjects [Carlyon et al., 2002] (see also Chapter 6). Secondly, only place cues are available in normal-hearing subjects for stimuli above 5000 Hz, because above this frequency the auditory nerve looses its capability to phase lock to the stimulus and consequently all temporal information related to the stimulus frequency is lost [Pickles, 1988]. In the acoustical model used in this chapter, place pitch cues will thus be changed by varying the center frequency of an (amplitude modulated) noise band whose spectrum is above 5000 Hz; and purely temporal pitch cues will be changed by varying the modulation frequency of this modulated noise band. In order to have a simpler naming, pitch mechanisms related to the site of activation (place pitch in CI subjects and spectral pitch in NH) are both called place pitch. Also purely temporal pitch in NH and temporal pitch in CI are both called temporal pitch in the rest of this chapter.

4.2 Methods

4.2.1 Subjects

Nine normal-hearing subjects (of which 3 females), with pure tone hearing thresholds below 20 dB HL between 500 and 8000 Hz, participated in this experiment. The subjects' age ranged from 22 years to 32 years. All subjects were

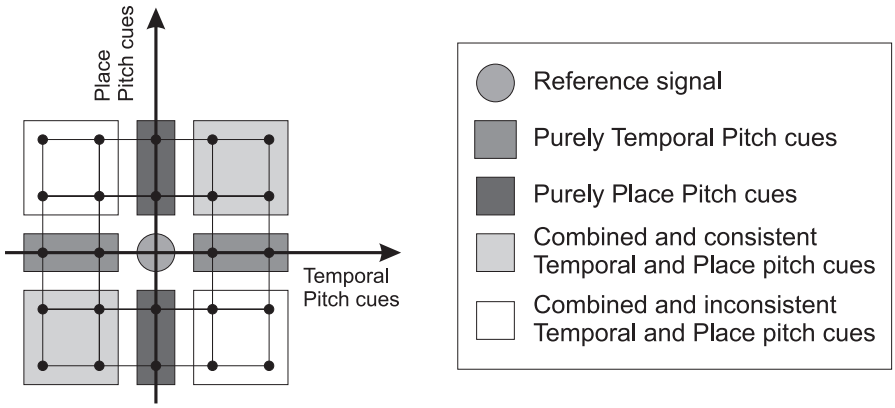


Figure 4.1: Graphical representation of the set of stimuli used in the experiment.

members of the staff of the Laboratory of Experimental ORL of the Katholieke Universiteit Leuven.

4.2.2 Stimuli

A set of stimuli was created which differ in place and/or temporal pitch cues from a reference stimulus. The stimulus space is shown in Figure 4.1. The large grey dot in the center of the figure represents the reference stimulus, and the smaller black dots represent the comparison stimuli. The reference stimulus is a noise-band with center frequency 6500 Hz and sinusoidally amplitude-modulated at 130 Hz. For the comparison stimuli, place pitch cues are varied by changing the center frequency of the noise band to 5700 Hz, 6100 Hz, 6500 Hz, 6900 Hz, or 7300 Hz. The center frequency of the noise bands was chosen so that no phase locking was possible at its center frequency but only to the modulation frequency. The temporal pitch cues were varied by increasing or decreasing the modulation rate with 7% or 14% compared to the modulation frequency of the reference stimulus. The resulting modulation rates of the comparison stimuli were thus approximately 112 Hz, 121 Hz, 130 Hz, 139 Hz, or 148 Hz.

The combination of the five different modulation rates and the five different center frequencies leads to 25 different comparison stimuli that can be divided into four groups.

1. A group of four comparison stimuli that only differ from the reference signal in temporal pitch cues or modulation rate (represented in Figure 4.1 by the two gray rectangles on the horizontal axis).
2. A group of four comparison stimuli that only differ from the reference signal in place pitch cues or center frequency of the noise band (represented in Figure 4.1 by the two dark gray rectangles on the vertical axis).
3. A group of eight comparison stimuli that differ in both place and temporal pitch cues from the reference signal, and where these pitch cue differences are consistent, i.e. both rising or both falling relative to the reference stimulus

(represented in Figure 4.1 by the pale gray squares in the upper-right, and lower-left corners).

4. A group of eight comparison stimuli that differ in both place and temporal pitch cues from the reference signal, but where these pitch cue differences are inconsistent, i.e. where one pitch mechanism indicates a rise in pitch and the other pitch mechanism indicates a fall in pitch (represented in Figure 4.1 by the white squares in the upper-left and bottom-right corners).

The noise bands were generated offline at a sampling frequency of 36 kHz and were 500 ms long including 30 ms cosine ramps at the start and end. The noise bands were created by filtering white Gaussian noise with a custom designed band-pass FIR filter: the CISIM filter (see chapter 6, section 6.2.2). The space constant of the exponential decay λ was 1 mm. The resulting filter has an approximately triangular shape with slopes of approximately 40 dB/octave. This filter was designed to roughly model the physiology of electrical stimulation through a cochlear implant. More details about this custom designed filter can be found in chapter 6.

4.2.3 Psychophysical procedure

The stimuli were presented in a 2 Interval – 2 Alternative Forced Choice (2I-2AFC) pitch ranking task with the method of constant stimuli. The subjects were presented with the reference stimulus and a comparison stimulus in random order and are asked to indicate the highest one in pitch. In order to enforce the subjects to use their intuitive sense of pitch, no feedback was given and no instruction was given to what kind of cue (other than pitch) the subjects had to pay attention. The two stimuli in a trial were separated by a 500 ms silent gap.

For each trial one of the 25 comparison stimuli was randomly selected (the reference stimulus was also compared to itself). The trials were presented in blocks consisting of 50 trials, because within each block every comparison stimulus was repeated twice. Each block was presented 25 times to each subject and the order of presentation within each block was randomized for every test run. Consequently, each condition (or comparison stimulus) was presented 50 times to each subject.

The stimuli were presented at an intensity randomly chosen between 58 dB SPL and 66 dB SPL to minimize the effect of loudness cues. In order to mask possible distortion products and to prevent off-frequency, a pink background noise was presented continuously at a level of 40dB SPL measured per one-third octave band listening. The stimuli and background noise were presented monaurally to the subject's best ear and subjects were seated in a double-walled soundproof booth.

The acoustical stimuli were routed via a desktop PC and a 24-bit PCI sound card (Lynx ONE, Lynx Studio Technology), to a mixer-amplifier (Eurorack MX1604A) and played to one earpiece of a Senheisser HD250 II-"linear" headphone. The APEX program (see chapter 2) controlled stimulus playback and response collection. The background noise was played from CD and mixed with the stimuli by the amplifier-mixer.

4.2.4 Data analysis

The number of times the comparison stimulus was indicated by the subject as higher in pitch than the reference stimulus was averaged per comparison stimulus and per subject over the 25 test runs. These proportions $p(\Delta f_m, \Delta f_c)$ of the trials were converted into *d'* values using:

$$d'(\Delta f_m, \Delta f_c) = \sqrt{2}.Z\left(p(\Delta f_m, \Delta f_c)\right), \tag{4.1}$$

where Z denotes the inverse cumulative normal distribution [Green and Swets, 1988]. Here Δf_m denotes the difference in modulation rate between the particular comparison stimulus and the reference stimulus, and is thus related to the amount of temporal pitch cues. Furthermore Δf_c denotes the difference in center frequency between the particular comparison stimulus and the reference stimulus, and is thus related to the amount of place pitch cues. Positive *d'* values indicate that on average the comparison stimulus was perceived higher in pitch than the reference stimulus. Negative *d'* values indicate that the comparison stimulus was perceived lower in pitch than the reference stimulus.

For comparison stimuli that were perfectly discriminated from the reference stimulus (i.e. either on every trial it was perceived higher or on every trial it was perceived lower than the reference stimulus, corresponding to proportions of 100% and 0%, respectively) the proportion was slightly adjusted to prevent infinitely large *d'* values. For the proportions corresponding to these comparison stimuli half a trial error was introduced as suggested by MacMillan and Creelman [1991].

4.3 Results

The obtained *d'* values for each comparison stimulus and for each subject are shown in Figure 4.2. Each panel shows the results for a particular subject and the layout of the panels is identical to the layout of Figure 4.1. Place pitch cues vary along the vertical axis and temporal pitch cues vary along the horizontal axis.

The results indicate that there is a large variance in sensitivity to both place pitch cues and temporal pitch cues. Just noticeable differences (JND) were estimated for each subject from the obtained *d'* values for the single cue conditions. The temporal pitch JND was estimated by the point that corresponds to *d'* = 1 for a linear curve fit to the results obtained with comparison stimuli varying only in modulation frequency compared to the reference stimulus. The place pitch JND was estimated by the point that corresponds to *d'* = 1 for a linear curve fit to the results obtained with comparison stimuli only varying in center frequency compared to the reference stimulus. This calculation of JND is approximately equivalent to estimating the 75% correct point on a cumulative normal distribution fitted to the proportions of correctly ranked trials as used in chapter 3. The obtained place and temporal JND values are shown in Figure 4.3.

The analysis of the estimated JNDs shows that there is a significant negative correlation between temporal pitch sensitivity and place pitch sensitivity (Spearman's rho = -0.783; p = 0.013). This means that some subjects were more

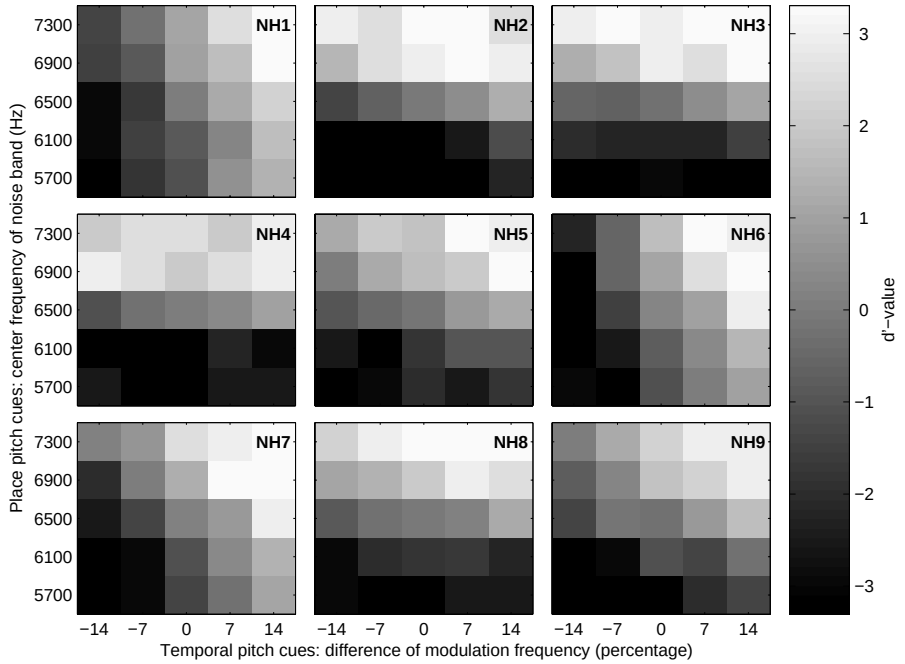


Figure 4.2: Results for each comparison stimulus and for each subject on the pitch ranking task. Positive d' values indicate that the comparison stimulus was perceived higher than the reference stimulus. Negative values indicate that the comparison stimulus was perceived lower in pitch than the reference stimulus.

sensitive to temporal pitch cues than others, but these subjects were less sensitive to place pitch cues than the others.

In order to statistically assess the discriminability of the stimuli (i.e. compare the sizes of the *d'* values) the absolute value of the measured *d'* values is used in the further statistical analysis. This allowed comparing falling and rising pitch changes as analogous cases under a different measurement condition.

In order to test for the combination of place and temporal pitch cues, the sizes of the *d'* values from conditions with consistent pitch cue combination were compared to the largest *d'* value of the corresponding single cue conditions using a repeated measures ANOVA. For instance, the *d'* values for the condition with center frequency 7300 Hz and modulation frequency 148 Hz were compared to the largest of the *d'* values from the condition with center frequency 6500 Hz and modulation frequency 148 Hz, and the condition with center frequency 7300 Hz and modulation frequency 130 Hz.

The analysis showed that the discriminability for combined cues was larger than the discriminability for any of the single cues [$F(1) = 11.865$; $p = 0.009$]. This means that the subjects were able to combine the place and the temporal pitch cues into a more salient combined pitch cue when the temporal and place pitch cues are consistent. Place and temporal pitch cues were constructively combined in a pitch ranking task when the cues are consistent. This is mathematically expressed as:

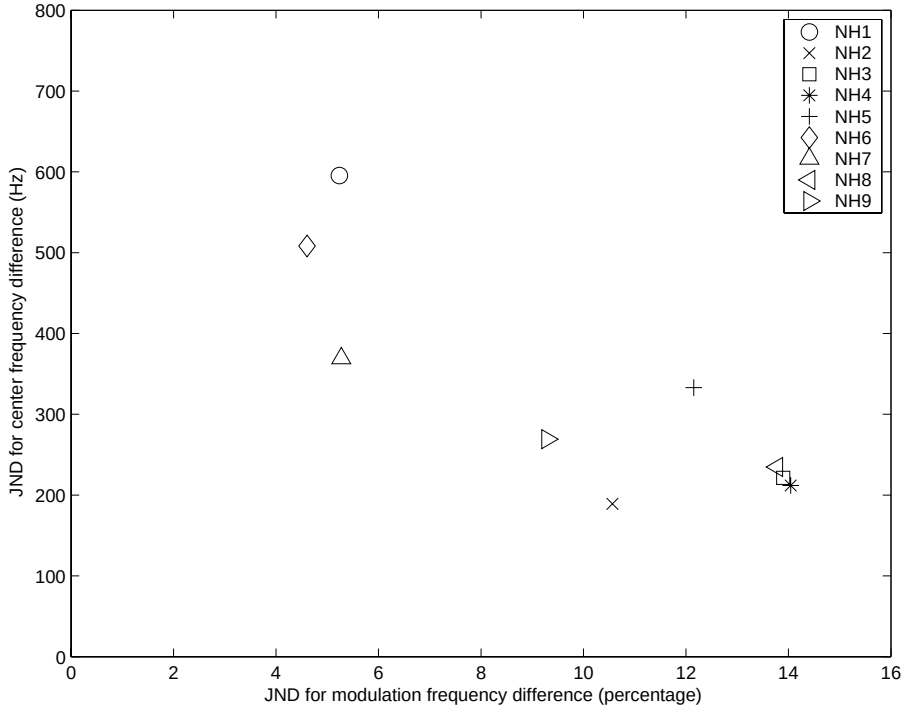


Figure 4.3: Scatter plot of sensitivity to place pitch cues and to temporal pitch cues for each subject. Smaller JNDs indicates better sensitivity to pitch cues.

$$\left|d'\left(\Delta f_m,\Delta f_c\right)\right| > \max\left[\left|d'\left(\Delta f_m,0\right)\right|,\left|d'\left(0,\Delta f_c\right)\right|\right] \text{ for } sign\left(\Delta f_m\right)=sign\left(\Delta f_c\right) \quad (4.2)$$

The size of the combined cue for inconsistent conditions was compared to the size of the largest corresponding single cue condition in a separate repeated measures ANOVA. This analysis showed that the size of the *d'* values for inconsistent-cue combination was smaller than the largest corresponding single cue *d'* value [$F(1) = 32.504$; $p = < 0.001$]. This can mathematically be expressed as:

$$\left|d'\left(\Delta f_m,\Delta f_c\right)\right| < \max\left[\left|d'\left(\Delta f_m,0\right)\right|,\left|d'\left(0,\Delta f_c\right)\right|\right] \text{ for } sign\left(\Delta f_m\right)=-sign\left(\Delta f_c\right) \quad (4.3)$$

The following results can be drawn. Firstly, the combination of inconsistent place and temporal cues results in lower salience than the largest of the corresponding single cues (purely temporal or purely spatial). Secondly, the relative sign of the single pitch cues is important because this effect is opposite to the effect found for conditions with consistent-pitch-cue combination.

On average, the size of the combined cue in the inconsistent condition was in between the size of the largest cue and the size of the smallest cues. This was because, in general, the size of the largest single cue was relatively large, and the size of the smallest single cue was relatively small, leading to a large range between the largest and smallest single cue. To test whether inconsistent cues interfere with each other, a selection of the inconsistent cues was made where the largest and the

smallest single cue did not differ more than 10%. In this way the range between the largest and smallest single cue was minimized. A t-test compared the measured *d'*-size for the selected inconsistent-cue combination to the *d'*-size for the smallest corresponding single cue. The size of the measured combined cues was smaller than the smallest corresponding single cue of those selected inconsistent cues ($p < 0.001$). This can be mathematically expressed as:

$$\left|d'\left(\Delta f_m, \Delta f_c\right)\right| < \min\left[\left|d'\left(\Delta f_m, 0\right)\right|, \left|d'\left(0, \Delta f_c\right)\right|\right] \tag{4.4}$$

for

$$sign\left(\Delta f_m\right) = -sign\left(\Delta f_c\right), \text{ and } 0.90 \le \frac{\left|\Delta f_m\right|}{\left|\Delta f_c\right|} \le 1.10$$

This means that inconsistent temporal and place cues when combined in a pitch ranking task, integrate destructively. The salience of the combined cues is lower than any of the single (purely temporal or purely spatial) cues when the sign of the temporal and place cues are opposite.

4.4 Discussion

The results showed that both increases in modulation frequency and increases in center frequency elicit rising pitch percepts in all subjects. The JNDs for changes in center frequency ranged from 190 Hz to 600 Hz among the nine subjects. Sensitivity of normal-hearing subjects to changes in center frequency depends on the center frequency and on the slopes of the filter [Gagne and Zurek, 1988]. Based upon the findings of Gagne and Zurek, the just noticeable difference for a second order resonator centered on 6.5 kHz with Q = 1.2 (which has a similar Q_{10dB} as the custom designed filter in the present experiments) would be around 470 Hz. The results of Gagne and Zurek thus agree with the results for center frequency discrimination in this experiment.

The JNDs for changes in modulation frequency ranged from 4.5% to 14% among subjects. This range of JNDs corresponds to the range of average JNDs from 5% to 8% for broadband amplitude modulated noise for modulation frequencies from 100 Hz to 200 Hz in normal-hearing subjects [Formby, 1985].

The measured *d'*-values for the cue-combination conditions can be expressed as a function of both *d'*-values of their respective single cue conditions. Figure 4.4 shows the $d'\left(\Delta f_m, \Delta f_c\right)$ for cue-combination conditions as a function of the measured purely temporal pitch cue $d'\left(\Delta f_m, 0\right)$, and the measured purely place pitch cue $d'\left(0, \Delta f_c\right)$ grouped over all subjects. Intermediate values were interpolated using the nearest neighbor method.

For the combination of multiple cues in general, a formula has been proposed previously to calculate the *d'* value of the combined cue from the single cue *d'* values (Eq. 1.15 in MacMillan and Creelman [1991]):

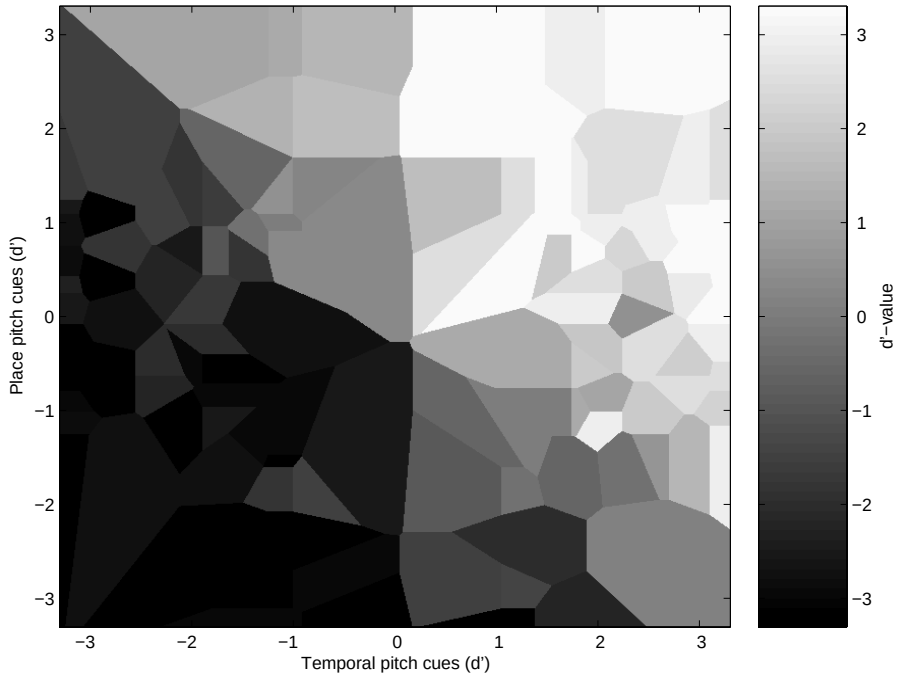


Figure 4.4: Nearest neighbor interpolation of group data of combined pitch cues as a function of temporal pitch cue and place pitch cue.

$$\left(d'_{combined}\right)^n = \sum_i \left(d'_i\right)^n \tag{4.5}$$

This general formula allows for modeling various modes of combination. For $n = 2$, the combination of the cues corresponds to the optimal integration of two independent cues with uncorrelated noises. For $n = 1$, the combination corresponds to the optimal integration of two cues that lie along the same dimension and of which the noises are fully correlated. For n between 1 and 2, the combination of the cues corresponds to the integration of cues with partially correlated dimensions or noises [Sek and Moore, 1999]. If n is smaller than 1, the combination is synergistic, i.e. the salience of the combined cue is larger than the sum of the salience of both single cues. If n is larger than 2, the combination is suboptimal. This means that the subjects fail to optimally combine the information of both cues. For n going to infinity, the model suggests no integration of the cues at all; i.e. the subjects are unable to combine both cues and just select the most salient one (the one with the largest d').

Place pitch cues and temporal pitch cues elicit sensations independent of each other [Tong et al., 1983]. According to signal detection theory [Green and Swets, 1988], the d' value for the optimal integration of independent cues corresponds to Eq. (4.5) with $n = 2$. For the integration of place pitch cues and temporal pitch cues the combined cue can thus be expressed as:

$$d'^2_{combined} = d'^2_{place} + d'^2_{temp} \tag{4.6}$$

This formula has been confirmed by McKay et al. [2000] using a detection-task in four cochlear implantees. Note that this formula discards the sign of the cues. However, in contrast to *d'* values obtained in a detection task (which can only be positive), the *d'* values obtained in a pitch ranking experiment can be both positive and negative. The results obtained in the present experiment also indicated that the sign of the cues can not be discarded and greatly influences the combined cue. Based upon our findings, Eq. (4.6) is adjusted to explicitly take into account the sign of the cues:

$$\text{sign}(d'_{combined})d'^{2}_{combined} = \text{sign}(d'_{place})d'^{2}_{place} + \text{sign}(d'_{temp})d'^{2}_{temp} \qquad (4.7)$$

This formula simplifies again to Eq. (4.6) for a detection-task as in McKay *et al.* [2000] because then only positive *d'* values are possible. The formulas for the integration of place and temporal pitch cues thus differ for a pitch-change-detection task and for a pitch-ranking task. This is similar to the concept of trading of interaural level differences and interaural timing differences in directional hearing. Hafter and Carrier [1972] have confirmed that observers are able to detect differences between diotic signals and dichotic signals which have been centered subjectively by opposing a time difference with an intensity difference. Hafter and Carrier also showed that the sign of the "intensity image" and the "time image" is not important for detecting the difference but the sign of both images obviously is important for the perceived lateralization.

Based upon Eqs. (4.6) and (4.7), predictions of the combined pitch cues were calculated for all conditions of combined pitch cues (consistent and inconsistent) and for each subject. The correlation between predicted and measured *d'* values was -0.10 (p = 0.22) and 0.95 ($p < 0.001$) for the two equations, respectively. This indicates that Eq. (4.6), which discarded the sign of the pitch cues, failed at predicting our measured results. Eq. (4.7) was able to closely predict the measured results and thus provides an accurate model for the integration of place and temporal pitch cues in a pitch ranking task.

The presented model implies that place and temporal pitch cues are able to cancel each other out in a pitch ranking task. This means that for two stimuli with place and temporal pitch cues in opposite directions (one falling and one rising) the discriminability can correspond to a *d'* value of zero. In other words, for this particular pair of stimuli the subject is unable to consistently tell which one of the two stimuli is perceived higher in pitch. However, this does not mean that the two stimuli are perceptually indifferent from each other. In fact, the two stimuli will elicit different sound sensations in the subject which can be discriminated as shown by McKay et al. [2000]. In short, when two stimuli differ in both place and temporal pitch cues but these cues are inconsistent, the subject may not be able to rank one stimulus as higher in pitch than the other, but clearly hears a difference between the two stimuli.

The fact that the *d'*-value approaches zero when the single pitch cues are of opposite sign and of approximately equal size, means that the subject is unable to tell which of the two stimuli sounds higher in pitch, and consequently has to guess. By chance,

the subject will indicate on average both stimuli an equal proportion of times as sounding higher in pitch.

However, another explanation may be possible and can not be excluded because the data are based upon the averaging of many trials. This assumption states that the subject applies some kind of cue switching when confronted with conflicting (pitch) cues. In this case, the subject sometimes bases his decision on the place pitch cue and the other times on the temporal pitch cue. The selected cue could depend on the relative size of the cue, attention, the previous trials or other factors. In case the subject applies cue-switching, the proportion of times the subject selects one cue over the other determines the *d'* value of the combined cue.

4.5 Conclusion

The experiment reported in this chapter measured the sensitivity of normal-hearing subjects to 'CI-simulated' place and temporal pitch cues in a pitch ranking task. The place and temporal pitch cues were either presented alone or combined. When both cues were presented together, the cues were consistent (either both rising or falling) or inconsistent (one rising and one falling).

The results show that subjects are able to combine both cues (place and temporal) in a more discriminable pitch sensation when the cues are consistent. The results also show that place and temporal pitch cues are able to cancel each other out when the cues indicate pitch changes in opposite directions (inconsistent pitch cues). The results are consistent with a model of the combination of pitch cues that explicitly takes into account the sign of the pitch cues.

Chapter 5

The effect of filter banks on fundamental frequency discrimination by cochlear implant subjects

J. Laneau, M. Moonen, and J. Wouters. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *Journal of the Acoustical Society of America,* 116(6):3606-3619, December 2004.

The speech processor of a cochlear implant system transforms the acoustic sound wave into a complex electrical stimulation pattern, activating many if not all available electrodes. In contrast, most studies investigating pitch perception in CI subjects, including the previous chapters of this thesis, used synthetic stimuli that do not originate from a real-world acoustic signal. These synthetic signals are most often rather simple compared to the signals presented to the CI-subjects by their clinical CI-systems in daily-life. In this chapter, pitch perception of harmonic signals processed through the sound-processor of the CI system is investigated. In particular

the effect of one part of the sound processor, the filter bank, will be assessed for the discrimination of the fundamental frequency (or pitch) of the harmonic input signals. In this way, pitch perception of real world signals by CI subjects is measured. Secondly, the performance of the subjects is compared over different filter banks for the sound processor in order to find an optimal sound processing scheme for pitch perception. A third goal of this chapter is to measure the relative contributions of the separate pitch mechanisms (place pitch and temporal pitch) in fundamental frequency (F0) discrimination by CI subjects.

Section 5.2 describes the set-up of the experiment of this chapter. The effect of the filter bank on F0 discrimination was examined in four Nucleus CI24 cochlear implant subjects for synthetic stylized vowel-like stimuli. The four tested filter banks differed in cut-off frequencies, amount of overlap between filters and shape of the filters. To asses the effects of temporal pitch cues on F0 discrimination, temporal fluctuations were removed above 10 Hz in one condition and above 200 Hz in another. The results, presented in section 5.3, indicate that F0 discrimination based upon place pitch cues is possible, but just noticeable differences exceed one octave or more depending on the filter bank used. Increasing the frequency resolution in the F0 range, improves the F0 discrimination based upon place pitch cues. The addition of temporal fluctuations up to 200 Hz significantly improves F0 discrimination. Just noticeable differences using both place and temporal pitch cues range from 6% to 60%. In the discussion of the results, section 5.4, it is shown that the results of F0 discrimination based upon place pitch agree with a model that compares the centroids of the electrical excitation pattern. The results also show that filter banks that do not resolve the higher harmonics, provided the best temporal pitch cues. Temporal pitch cues are clearest when the fluctuation on all channels is at F0 and preferably in phase. The chapter concludes that no filter bank tested is optimal for pitch perception for place and temporal pitch mechanisms simultaneously and alternative sound processing strategies are required to improve pitch perception by CI subjects.

5.1 Introduction

It has been shown that both temporal and place pitch cues enable CI-recipients to some degree to perceive F0 differences of synthetic harmonic sounds with currently used sound processors [Geurts and Wouters, 2001], and in normal-hearing subjects with acoustic models of CI-processors [Green et al., 2002]. Geurts and Wouters [2001] measured the discrimination of F0 of synthetic steady-state vowels in subjects implanted with the 8-channel Laura implant. Their results indicated that even without temporal pitch cues, only preserving place pitch cues, some discrimination of F0 was possible although the JNDs were very large. Without temporal pitch cues the JNDs were on average around 50% but varied over conditions and subjects from 5% to more than 100%. With the inclusion of temporal information, the JNDs decreased significantly, as would be expected. Green et al. [2002] examined the discrimination of CI simulations of frequency glides in sawtooth waves and synthetic diphthongs in normal-hearing subjects. They studied both place and temporal pitch in modulated noise band vocoder CI-simulations, and concluded that both pitch mechanisms contribute to the discrimination of F0.

Subjects were able to integrate both cues as the inclusion of both pitch cues produced the highest discriminative performance. However, when the original sounds were diphthongs, the inherent variation of formant structures perturbed the place pitch cue.

The current sound-processing in cochlear implants has a negative effect on F0 discrimination since CI-subjects appear to be able to estimate musical intervals more accurately for pulse trains than for real musical sounds processed by their speech processors [Pijl, 1997]. Part of this limited F0 discrimination ability in CI recipients using their speech processors may be due to the limited transmission of the fine temporal information in current speech processors, since most current speech processors only extract the slowly varying envelope and use this to modulate a constant rate pulse train [Smith et al., 2002].

Because of this limited use of fine temporal pitch cues in current CI-sound processors, previous studies have tried to improve the transmission of temporal pitch cues in order to improve the F0-discrimination in cochlear implants [Jones et al., 1995; Geurts and Wouters, 2001; Barry et al., 2002]. However, none of these studies resulted in a significant improvement of F0 discrimination. Explicit presentation of the fundamental frequency by stimulating the most apical channel with a pulse train at the rate of F0 did not improve the perception of suprasegmental cues such as intonation [Jones et al., 1995]. Barry et al. [2002] assessed the effect of pulse rate on the perception of tones in Cantonese speaking children by comparing a group of children using the SPEAK strategy to a group of children using the higher rate ACE strategy. The hypothesis was that higher rates better preserve the temporal code [Wilson, 1997; Rubinstein et al., 1999] but the authors found no significant differences between the strategies. Indeed, the group of children with the ACE strategy performed slightly (but non-significantly) worse than the group of SPEAK users. Geurts and Wouters [2001] developed a variation to the CIS sound processing scheme that enhanced the modulation depth of the beating of the output of the filters, and also did not find a significant difference when compared to the standard CIS in the discrimination of F0 of synthetic vowels.

In a recent study, Geurts and Wouters [2004] showed that a new filter bank can improve the place pitch cues for F0 discrimination of synthetic vowels. Their newly proposed filter bank had more resolution in the lower frequencies and more filter overlap between different channels compared to the reference filter bank in that study. An important improvement in F0 discrimination based only upon place pitch cues was obtained but it remains, however, unclear whether this is due to the increased frequency resolution, the smooth transition between filters, or the increased filter overlap. With the inclusion of temporal cues, the improvement was small or even absent.

The present study investigates which aspect of the filter bank (cut-off frequency distribution, filter shape, or filter overlap) has the strongest effect on F0 discrimination for both place pitch cues and temporal pitch cues. It also investigates which physical properties of the stimuli are related to the place pitch cues that facilitate F0 discrimination, and which lead to clearer temporal pitch cues.

The effect of four filter banks on the voice-range pitch discrimination is compared in cochlear implant users. In order to separate the effects of place pitch cues and temporal pitch cues on complex pitch perception, two experimental conditions were created. In the first condition, temporal pitch cues were excluded by low-pass filtering the stimulation envelopes at 10 Hz, and consequently only place pitch cues were present. In the second condition both place and temporal pitch cues (up to approximately 200 Hz) are present.

Performance on an F0-discrimination task was measured as a function of relative difference of fundamental frequency, using a constant-stimuli 2I2AFC pitch-ranking experiment. Four Nucleus CI24 cochlear implant subjects took part. The harmonic signals were synthetic stylized vowel-like single formant stimuli.

5.2 Methods

5.2.1 Subjects

Four experienced postlingually deafened users of the Nucleus CI24 cochlear implant participated in this study. All subjects performed all tests and were paid for their collaboration. The subjects also participated in the experiment of Chapter 3. Some relevant details about the subjects can be found in Table 3.1. Subjects CI1 and CI2 are implanted with the straight version of the Nucleus electrode array CI24R(ST) Nucleus24k and use a SPEAK strategy in daily life. Subjects CI3 and CI4 are implanted with the peri-modiolar version of the Nucleus electrode array CI24R(CS) Nucleus24 Contour and use an ACE strategy (900 pps per channel) in daily life. All subjects are relatively good performers with their implants in daily life.

5.2.2 Stimuli

Original stimuli before sound-processing were synthetic harmonic complexes produced by the excitation of a single formant resonator by a periodic signal, and resemble highly stylized vowels. Figure 5.1 shows the spectrum of two representative stimuli. The generation of the single-formant stimuli (SFS) was based upon the Klatt speech-synthesizer [Klatt, 1980]. The stimuli were generated by filtering a 500ms pulse train sampled at 16 kHz and ramped on and off with a raised cosine over 10 ms with a cascade of two filters.

The first filter models the voice source and its output is similar to the typical glottal volume velocity waveform. It is implemented as a second order low-pass filter with a cut-off frequency of 50 Hz [Klatt, 1980]. The second filter is a resonator that shapes the single formant of the stimuli. It is implemented as a second order band-pass filter of which the transfer function is defined by

$$H(z)=\frac{a}{1-bz^{-1}-cz^{-2}} \quad with \quad \begin{aligned} c &= -\exp\left(-2\pi B_w T_s\right) \\ b &= 2\exp\left(-\pi B_w T_s\right)\cos\left(2\pi F_r T_s\right), \\ a &= 1-c-b \end{aligned} \qquad (5.1)$$

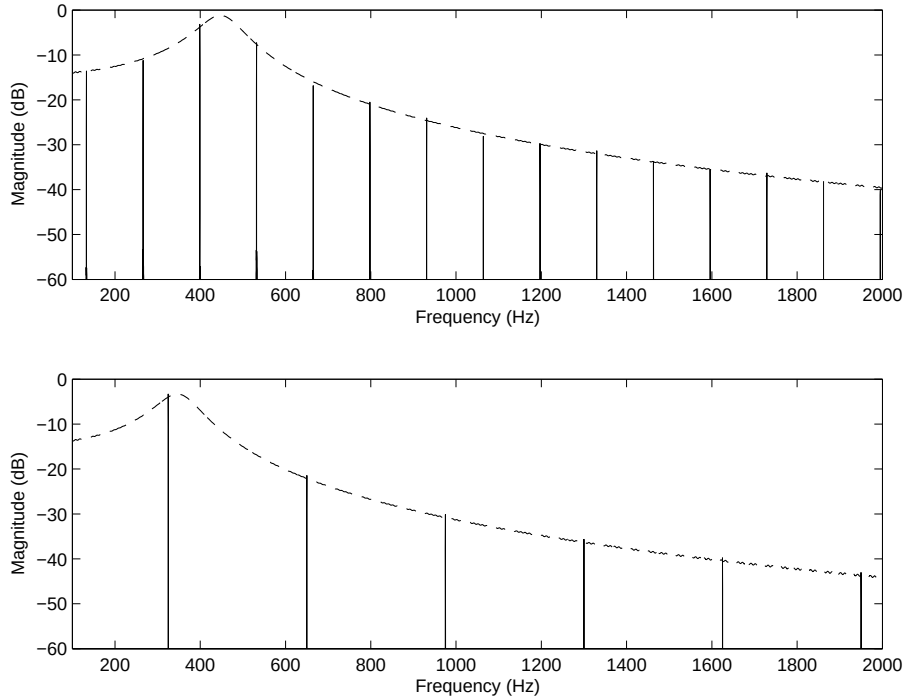


Figure 5.1: The spectrum of two representative stimuli used in the present study. The upper panel shows the spectrum of a single formant stimulus with formant frequency 450 Hz and fundamental frequency 133 Hz. The lower panel shows the spectrum of a single formant stimulus with formant frequency 350 Hz and fundamental frequency 325 Hz. The dotted line in each panel shows the combined frequency response of the two filters that shape the spectrum (see text). The plot of the frequency response is limited up to 2000 Hz for clarity.

where T_s is the sampling period (62.5 μs) and F_r and B_w are the filters center frequency and bandwidth, respectively [Klatt, 1980].

The fundamental frequency of the stimuli used in the present study ranged from 133 Hz up to 558 Hz. The center frequencies of the band-pass filter, or formant frequencies (F1), were 300, 350, 400, 450, or 500 Hz, and the bandwidth of the formant resonator was fixed at 100 Hz for all stimuli. The stimuli were first equalized in RMS and then scaled with a common factor so that most of the signals nearly optimally used the input dynamic range of the compression function (see below: Eq. (5.17)) but none of the signals was clipped.

5.2.3 Sound processing

Four different filter banks were used to process the stimuli: ACE, GTF, BUT, and GTFW. The frequency response of the four filter banks is depicted in Figure 5.2. The filter banks differ in spectral resolution for the low frequencies and in the shape of the individual filters. All filter banks have 22 bands corresponding to the number of channels in the Cochlear Nucleus CI24 using monopolar stimulation mode. One

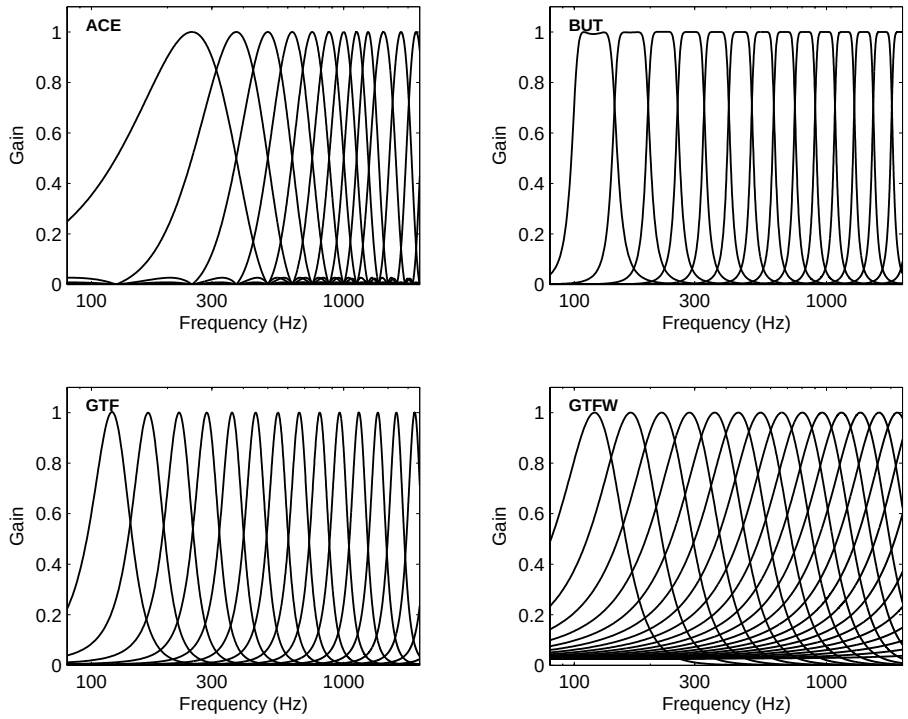


Figure 5.2: The frequency response of the different filter banks examined in the present study. Because the stimuli in the present study contained virtually no energy in higher frequencies, the frequency response is represented up to 2000 Hz for clarity. The frequency response for the ACE filter bank is obtained for pure tone signals processed with envelope extraction. The frequency response of the GTF, BUT and GTFW are obtained from the impulse response without envelope extraction.

filter bank (ACE) is at present most commonly used clinically for Nucleus CI24, and is included in the present experiment as a reference for current performance of cochlear implant subjects. The three other filter banks are specifically designed for this experiment and resemble more closely the frequency analysis of the normal human ear. All three filter banks have more resolution in the low frequencies compared to the ACE filter bank. The GTF filter bank is based upon a model for the filtering of the normal ear: the gammatone filter [Patterson et al., 1995]. The BUT (Butterworth) and GTFW (Wide Gammatone) are derived from this filter bank by changing the filters' shape and bandwidth, respectively.

5.2.3.1 The ACE filter bank

The first filter bank (ACE) is the filter bank of the standard commercial speech processing strategy (Cochlear Corp: ACE) of the Cochlear SPrint speech processor implemented in Matlab [Cochlear Ltd, 2002]. In summary, the filter bank analyzes the input signal with a 128-point FFT and the Hilbert envelope is extracted for each frequency bin of the spectrum. The extracted envelopes of all 65 frequency bin are

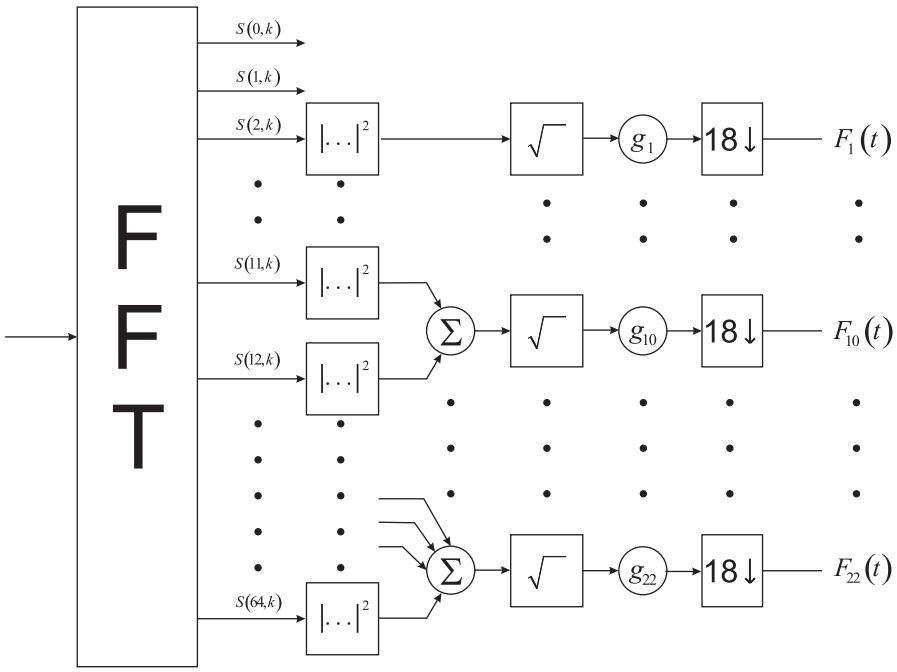


Figure 5.3: Overview of the processing blocks in the ACE filter bank.

then summed into 22 separate channels. The block diagram of the processing stages in the filter bank is given in Figure 5.3. The details of the implementation of the filter bank are discussed below.

The filter bank reads in the input signal $x[k]$, sampled at 16 kHz, and stores it into buffers of 128 samples ($L=128$) with 127 samples of overlap, or a step of only 1 sample per buffer. This leads to an update rate F_{update} of 16000 buffers per second, equal to the sampling rate. This differs from the actual implementation of the ACE filter bank in the clinical SPrint speech processor. In the SPrint processor the filterbank output is downsampled 21 times in order to reduce the computational load, leading to an update rate of approximately 762 Hz. However, in the present study the computational load is unimportant because all stimuli were processed off-line, and therefore no down-sampling was applied in the filter bank.

The spectrum of each buffer $S(q,k)$ at time instant k is obtained through an FFT after applying a Hanning window $w[k]$:

$$S(q,k)=\sum_{l=1}^{L} x[l]w[l]e^{-j\frac{2\pi}{L}lq} \quad ,q=0...L-1, \tag{5.2}$$

with

$$w[k] = 0.5\left(1.0 - \cos\left(\frac{2\pi k}{L}\right)\right) \quad , k = 0...L-1 . \tag{5.3}$$

The resulting spacing between adjacent frequency bins of the spectrum (q and $q+1$) is 125 Hz and the bandwidth of the main lobe of each frequency bin is 500 Hz [Oppenheim and Schafer, 1999].

Table 5.1: The assignment of the frequency bins to the 22 channels for the ACE filter bank.

Channel Number	Number of bins	Center freq. low bin (Hz)	Center freq. high bin (Hz)
1	1	250	250
2	1	375	375
3	1	500	500
4	1	625	625
5	1	750	750
6	1	875	875
7	1	1000	1000
8	1	1125	1125
9	1	1250	1250
10	2	1375	1500
11	2	1625	1750
12	2	1875	2000
13	2	2125	2250
14	3	2375	2625
15	3	2750	3000
16	4	3125	3500
17	4	3625	4000
18	5	4125	4625
19	5	4750	5250
20	6	5375	6000
21	7	6125	6875
22	8	7000	8000

For each frequency bin the envelope is estimated, based upon the definition of the envelope in the Hilbert Transform, by calculating the magnitude of the complex spectrum $|S(q,k)|$. Because the bandwidth of each frequency-bin is limited up to 500 Hz, the extracted envelopes are limited to 250 Hz.

In the next stage the 65 frequency bins are combined into 22 channels by summing the squared magnitude of a number of frequency bins for each channel. The lowest 2 frequency bins of the spectrum (including the DC bin) are discarded, because most

speech signals do not contain information critical for speech understanding below 200 Hz. Consequently, the filter bank spectrum ranges from 250 Hz up to 8000 Hz. The remaining 63 frequency bins are combined into 22 channels such that the bandwidths of these channels are distributed uniformly below 1000 Hz and logarithmically above 1000 Hz (roughly mimicking the physiology of the normal ear, see chapter 1). The assignment of the bins to each channel is shown in Table 5.1.

The amplitude of each channel is defined by the sum of the powers of the assigned bins:

$$F_n(k) = g_n \sqrt{\sum_q |S(q,k)|^2} \quad ,n = 1...22 \tag{5.4}$$

Where $F_n(k)$ is the amplitude of filter bank output channel *n* at time instant *k*, and $S(q,k)$ is the q^{th} frequency bin of the FFT of the buffer at time instant *k*.

The weights g_n are inserted to equalize the maximum gain of all channels and to scale the output so that its dynamic range is exactly from 0 to 1, taking into account that the input signal is limited between -1 and 1. The gains differ across channels because for channels consisting of multiple frequency-bins the spectral overlap of adjacent frequency-bins increases the maximum gain. The weights g_n are defined by:

$$g_n = \begin{cases} \dfrac{1}{|W(0)|} & for \quad channels\ with\ 1\ bin \\ \dfrac{1}{\sqrt{2}|W(0.5)|} & for \quad channels\ with\ 2\ bins \\ \dfrac{1}{\sqrt{|W(0)| + 2|W(1)|}} & for \quad channels\ with\ more\ than\ 2\ bins \end{cases} \tag{5.5}$$

where $|W(q)|$ is the magnitude of the q^{th} frequency-bin (where $q = 0$ is DC) of the FFT-spectrum of the Hanning window:

$$W(q) = \sum_{l=0}^{L} w[l] e^{-j\frac{2\pi}{L}ql} \quad ,q = 0...L\,. \tag{5.6}$$

For the calculation of $W(0.5)$ the window is zero-padded to twice its length before applying the FFT to increase the frequency resolution [Oppenheim and Schafer, 1999].

Finally the output of each channel is resampled to 900 samples per second (equal to the stimulation rate in the CI per channel). This is approximately equivalent to a 18-fold downsampling.

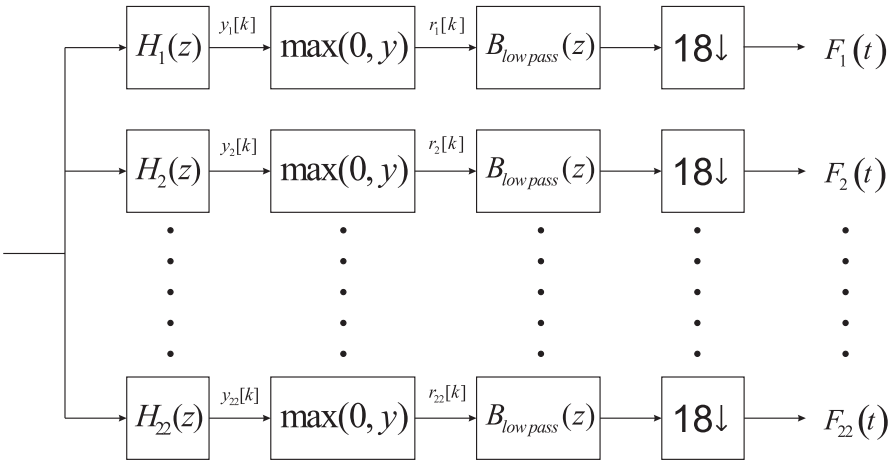


Figure 5.4: Overview of processing blocks in the three IIR filter banks (GTF, BUT, and GTFW).

The ACE filter bank is thus a highly non-linear filter bank, because of the envelope extraction. For a pure tone processed through the filter bank the output will be stationary in time (a DC). For an amplitude modulated tone, the output of the filter channel will be modulated at the same modulation rate as long as the modulation rate is relatively low (below 250 Hz). Above this frequency the 'side bands' of the amplitude modulated signal fall into separate frequency bins, and the different components of the signal are 'resolved'. In this case only pure tones are transmitted in each frequency bin and the output is stationary in time.

5.2.3.2 The IIR filter banks

The three IIR filter banks (GTF, BUT, and GTFW) are based upon the implementation of the CIS algorithm [Wilson et al., 1991]. The general block scheme is given in Figure 5.4. The first stage splits up the input signal $x[k]$ into 22 channels using a filter bank of 22 IIR band-pass filters. The GTF, BUT and GTFW filter banks all consist of 4th order IIR band-pass filters.

The output $y_n[k]$ of filter n of the filter bank is thus given by:

$$y_n[k] = \sum_{m=1}^{8} a_{m,n} y_n[k-m] + \sum_{m=0}^{8} b_{m,n} x[k-m] \quad , n = 1...22 \tag{5.7}$$

The coefficients $a_{m,n}$ and $b_{m,n}$ are determined by the filter bank.

The filters in the GTF filter bank were 4th order gammatone filters using Slaney's implementation [Slaney, 1993]. The center frequencies of the gammatone filters $F_c^{GTF}(n)$ were set corresponding to the middle of 22 equally long sections along the cochlea $x\left(F_C^{GTF}(n)\right)$ ranging from 100 Hz to 8000 Hz in total.

$$x\left(F_c^{GTF}(n)\right) = x(100) + (n-0.5)\frac{x(8000)-x(100)}{22} \quad ,n = 1...22 \tag{5.8}$$

The transformation from frequency to distance along the cochlea $x(f)$ was calculated by Greenwood's formula for a cochlear length of 35 mm and α = 0.06 [Greenwood, 1990]. The bandwidth of the filter was set to the equivalent rectangular bandwidth (ERB) of the auditory filter corresponding to its center frequency [Glasberg and Moore, 1990]:

$$BW_{GTF}(n) = 1.019 * 24.7\left(\frac{4.37}{1000}F_c^{GTF}(n)+1\right) \quad ,n = 1...22\,. \tag{5.9}$$

The discrete-time Gammatone filters are derived from continuous-time Gammatone filters. The frequency response of the continuous Gammatone filters is given by [Slaney, 1993]:

$$\begin{aligned} GTF_n(s) = \; & \frac{1}{\left(s^2 + 2BW_{GTF}(n)s + BW_{GTF}(n)^2 + \left(2\pi F_c^{GTF}(n)\right)\right)^4} \\ & .\left\{\left(s-\left(-BW_{GTF}(n) - j2\pi F_c^{GTF}(n)\right)\right)^4\right. \\ & \left.+\left(s-\left(-BW_{GTF}(n) + j2\pi F_c^{GTF}(n)\right)\right)^4\right\} \end{aligned} \tag{5.10}$$

The coefficients $a_{m,n}$ and $b_{m,n}$ for the discrete time Gammatone filter are derived from this continuous-time filter using the bilinear transform with pre-warping around the center frequency of the channel $F_c^{GTF}(n)$ [Oppenheim and Schafer, 1999].

The BUT filter bank is similar to the GTF filter bank, however the shape of the filters is more rectangular (a flat frequency response in the pass band and 24dB/octave slopes). The BUT filter bank consists of 22 4th order Butterworth filters of which the cut-off frequencies were equidistantly spaced along the length of the cochlea with cut-off frequencies ranging from 100 Hz to 7800 Hz in total. The highest cut-off frequency was set slightly below the Nyquist frequency to avoid numerical problems. The coefficients $a_{m,n}$ and $b_{m,n}$ for the BUT filter bank are determined by the 'butter' function in Matlab. The 'butter' command in Matlab calculates the continuous-time filter and transform it into a discrete-time filter using the bilinear transform. However, the transform is adjusted so that the magnitude of the frequency response is exactly 3dB attenuation also for the discrete-time filter. Both the GTF and the BUT filter bank resolve one or two of the lowest harmonics for most of the stimuli in the present study.

The GTFW filter bank is based upon the GTF filter bank but the bandwidth of the filters is increased so that all overtones are unresolved. The center frequencies of the Gammatone filters of the GTFW filter bank have the same center frequencies as the Gammatone filters of the GTF filter bank.

$$F_c^{GTFW}(n) = F_c^{GTF}(n) \quad n = 1...22 \tag{5.11}$$

Table 5.2: Cut-off frequencies for the IIR filter banks. The cut-off frequencies are determined by the frequency where the mangitude of the frequency response is 3 dB below the maximum gain of the specific channel.

	GTF		BUT		GTFW	
Channel number	Lower cut-off freq. (Hz)	Upper cut-off freq. (Hz)	Lower cut-off freq. (Hz)	Upper cut-off freq. (Hz)	Lower cut-off freq. (Hz)	Upper cut-off freq. (Hz)
1	104	137	101	143	94	147
2	149	187	145	196	130	204
3	201	244	196	257	173	271
4	262	311	257	327	223	349
5	333	389	327	410	281	440
6	415	480	410	506	348	545
7	512	585	506	618	427	668
8	624	708	618	749	519	810
9	755	852	749	902	626	976
10	907	1018	902	1080	751	1168
11	1085	1212	1080	1288	898	1390
12	1292	1438	1289	1532	1071	1647
13	1535	1699	1532	1815	1273	1942
14	1818	2004	1816	2147	1512	2281
15	2149	2356	2147	2533	1796	2669
16	2536	2765	2533	2984	2134	3109
17	2991	3237	2985	3511	2541	3606
18	3526	3782	3511	4126	3036	4162
19	4156	4411	4126	4844	3645	4781
20	4899	5133	4844	5681	4403	5466
21	5777	5965	5681	6659	5358	6225
22	6813	6921	6659	7800	6560	7068

The bandwidth of each Gammatone filter was set to half the center frequency of the respective channel:

$$BW_{GTFW}(n) = \frac{F_c^{GTFW}(n)}{2} \quad , n = 1...22 \tag{5.12}$$

The increased bandwidth (relative to the GTF filter bank) caused the output of every filter for a harmonic sound to beat at the fundamental frequency because the channel contained only the fundamental or contained at least two overtones. The coefficients

$a_{m,n}$ and $b_{m,n}$ for the GTFW filter bank were derived identically to the GTF filter bank using (5.10) and the bilinear transform; except that the bandwidth was set at half the filter's center frequency as stated.

The resulting cut-off frequencies of all the filters of the three IIR filter banks (GTF, BUT, and GTFW) are presented in Table 5.2.

The envelope extraction for the IIR filter banks consists of a half-wave rectifier and a low-pass filter for each channel. The half-wave rectifier is chosen because it mimics the transduction of the hair cells in the normal ear. Another benefit of the half-wave rectifier over a full wave rectifier, as used in the CIS scheme [Wilson et al., 1991], is that with the half-wave rectifier the envelope fluctuation are at the same rate as the modulation frequency of the input signal while with a full wave rectifier the envelope fluctuation might be at a rate double the modulation frequency of the input signal. The output of the rectifier $r_n[k]$ is given by:

$$r_n[k] = \begin{cases} y_n[k] & if \quad y_n[k] \geq 0 \\ 0 & if \quad y_n[k] < 0 \end{cases} \quad , n = 1...22 \tag{5.13}$$

or equivalently

$$r_n[k] = \max\left(0, y_n[k]\right) \quad , n = 1...22 \, . \tag{5.14}$$

After the signal is rectified it is low-pass filtered to extract the slowly varying envelope. The low-pass filter $B_{low\,pass}(z)$ is a 4th order Butterworth filter. This low-pass filter is identical for all channels of the filter banks ACE, BUT, and GTFW, except that the cut-off frequency was either 200 Hz or 10 Hz (see below). The coefficients of the Butterworth filter c_m and d_m were determined by the 'butter' command in Matlab.

$$F_n[k] = \sum_{m=1}^{4} c_m F[k-m] + \sum_{m=0}^{4} d_m r_n[k-m] \quad , n = 1...22 \tag{5.15}$$

Then finally the low-pass filtered envelope $F_n[k]$ is resampled to a rate of 900 samples per second $F_n(t)$, which corresponds roughly to an 18 fold down sampling.

5.2.3.3 Envelope extraction with or without temporal pitch cues

In order to separate the effects of temporal and place pitch cues, the envelopes of the filter outputs were obtained in two different ways. In one condition, including both temporal and place pitch cues, the envelope contained modulations up to 200 Hz. In the other condition, including place pitch cues but without temporal pitch cues, the envelope only contained modulations up to 10 Hz. The deletion of modulations at frequencies above 10 Hz removes all fluctuations that generate temporal pitch percepts, although listeners can still understand speech, at least when presented in a quiet background [Drullman et al., 1994; Shannon et al., 1995].

In the ACE filter bank, the envelope is extracted by taking the magnitude of the complex-valued output, as described above. The bandwidth of the frequency bins of

the FFT-filter bank limits the maximal modulation frequency after this envelope extractor to approximately 200 Hz. In the condition without temporal pitch cues these filter bank outputs $F_n(n)$ were, in addition to the standard ACE processing, low-pass filtered with a 4th order Butterworth filter at 10 Hz. This low-pass filter was inserted just before the down sampling.

The envelopes for the GTF, BUT and GTFW were extracted using a half-wave rectifier (Eq. (5.14)) and a 4th order Butterworth low-pass filter $B_{low\,pass}(z)$. The cut-off frequency of the low-pass filter was set at 200 Hz when temporal pitch cues were transmitted and at 10 Hz when temporal pitch cues were removed.

5.2.3.4 Further processing

The subsequent processing was identical for all filter banks. The filter bank outputs $F_n(t)$ were sampled at 900 Hz. Then, only the 8 most intense channels per time slot were selected for stimulation. This is mathematically expressed as:

$$A_n(t) = \begin{cases} F_n(t) & if \quad F_n(t) \geq \max\limits_n 8\left(F_n(t)\right) \\ 0 & if \quad F_n(t) < \max\limits_n 8\left(F_n(t)\right) \end{cases}, n = 1...22\,, \tag{5.16}$$

where $\max\limits_n 8\left(F_n(t)\right)$ is the amplitude of the 8th largest channel at time instant t.

Finally, the amplitude of the signals is compressed using the standard compression function of the Nucleus Matlab Toolbox [Cochlear Ltd, 2002], in order to map the acoustic input level onto the subjects' electrical dynamic range. The compression function is defined by:

$$C_n(t) = \begin{cases} 0 & for \quad 0 \leq A_n(t) < 0.0156 \\ \log\left(1 + 416.2\dfrac{A_n(t) - 0.0156}{0.5703}\right) \Big/ 6.033 & for \quad 0.0156 \leq A_n(t) < 0.5859 \\ 1 & for \quad 0.5859 \leq A_n(t) \leq 1 \end{cases} \tag{5.17}$$

In Eq. (5.17), $A_n(t)$ is the amplitude of the extracted envelopes for each channel and $C_n(t)$ is the compressed amplitude expressed in proportion of the dynamic range expressed in clinical current units of the subject for that channel. The clinical current units $cl_n(t)$ for each channel of the stimuli are mapped between the T and C-levels of each subject:

$$cl_n(t) = \mathrm{T_n} + (\mathrm{C_n} - \mathrm{T_n})\, C_n(t) \quad , n = 1...22 \tag{5.18}$$

Clinical current units in the Nucleus CI24 device are exponentially related to the current amplitude measured in micro-amperes $I_n(t)$:

$$I_n(t) = 10 \cdot 175^{\frac{cl_n(t)}{255}} \quad , n = 1...22 \tag{5.19}$$

The obtained envelopes were used to modulate pulse trains of 900 pulses per second with biphasic pulses of 25 μs phase width and 8 μs inter phase gap. The pulses were delivered in monopolar mode with both external return electrodes active (MP1+2).

5.2.4 Psychophysical procedure

Fundamental frequency discrimination was measured in a two-interval, two-alternative forced choice (2I-2AFC) procedure using the method of constant stimuli. The subjects were presented with two stimuli and were instructed to indicate which of the two intervals contained the stimulus with higher pitch, while ignoring the loudness cues. The intervals were separated by a 500 ms silent gap and no feedback was provided, in order to force the subjects to base their judgements on a perceptual dimension that corresponded to their intuitive sense of pitch. This further reduced the likelihood of any residual loudness cues affecting their judgements.

Both stimuli in a trial had the same formant frequency. The F0 of one stimulus in each 2I-2AFC trial was equal to the reference frequency, and that of the other was set at a higher frequency. This reference frequency was 133 or 165 Hz. These frequencies lie in the range of normal adult voice, and 133 Hz is set near the first crossover between filters for the IIR-filter banks, and 165 Hz is set near the center of the second filter of the IIR-filter banks. The F0-frequencies of the 8 comparison stimuli were 140, 150, 165, 180, 210, 250, 325, and 450 Hz for the reference stimuli with an F0 of 133 Hz. For the reference stimuli with an F0 of 165 Hz, the 8 comparison stimuli had F0 frequencies of 174, 186, 205, 223, 261, 310, 403, and 558 Hz. The relative differences between F0 are the same for both reference F0 conditions and range from 5.2% up to 238%. The formant frequency was kept constant within trials because stimuli with the same F0 but with their main energy in different frequency regions can lead to differences in pitch [Ohgushi, 1978].

The trials were presented in blocks. All sixteen conditions resulting from the combinations of type of envelope extraction (with or without temporal pitch cues), filter bank (ACE, GTF, BUT or GTFW) and reference F0 (133 or 165Hz), were presented into separate blocks. Each block contained trials with all eight comparison frequencies for all five formant frequencies. These 40 different trials were repeated twice, leading to 80 trials per block. Different formant frequencies were included per block in order to prevent subjects from identifying and learning the reference stimulus based upon other cues. Every block was presented twice to every subject. Care was taken that blocks processed with the ACE, GTF and BUT filter bank were presented in random order. The blocks processed with the GTFW filter bank were presented at the end of the experiment.

To reduce the influence of loudness cues, amplitude was roved by multiplying the amplitude of all stimuli with a random factor between 0.85 and 1.1, with amplitude expressed in clinical current units and relative to the dynamic range. An exception was made for subject CI4 in the condition of the ACE filter bank without temporal pitch cues and the reference F0 of 165 Hz, where the subject judged the loudness of the stimuli as uncomfortably loud using the default amplitude roving. In this one case, the amplitude was roved with a factor between 0.77 and 1.

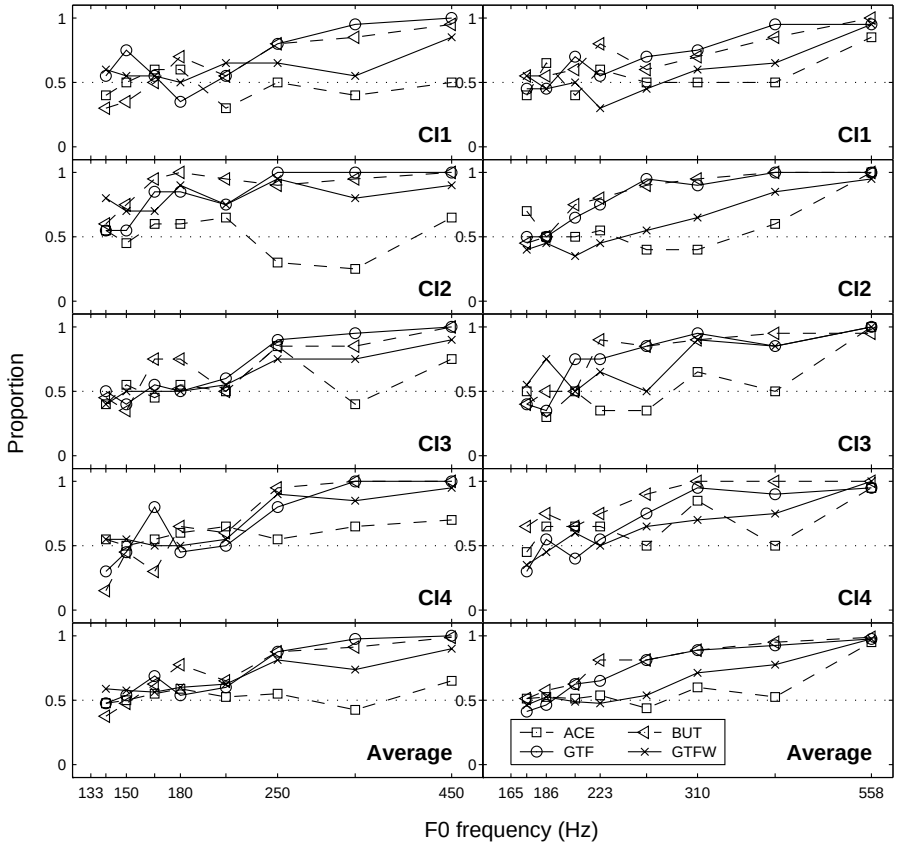


Figure 5.5: Results without temporal pitch cues: only place pitch cues. The average proportion of correctly pitch ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations above 10 Hz filtered out. The fundamental frequencies of the signals are plotted on the horizontal axis on a log-scale. The left and right columns of panels depict the results for the condition where the reference F0 was 133 Hz and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

5.3 Results

The proportion of times the comparison stimulus was judged higher than the reference stimulus, pooled over all formant frequencies, is shown in the absence and presence of temporal pitch cues in Figure 5.5 and Figure 5.6, respectively. The top four rows of panels in each figure show the results for a single subject, and the mean data are shown in the bottom, fifth row of panels. The left column in every figure shows the results for the condition with a reference F0 of 133 Hz and the right column shows the results for the condition with a reference F0 of 165 Hz. The

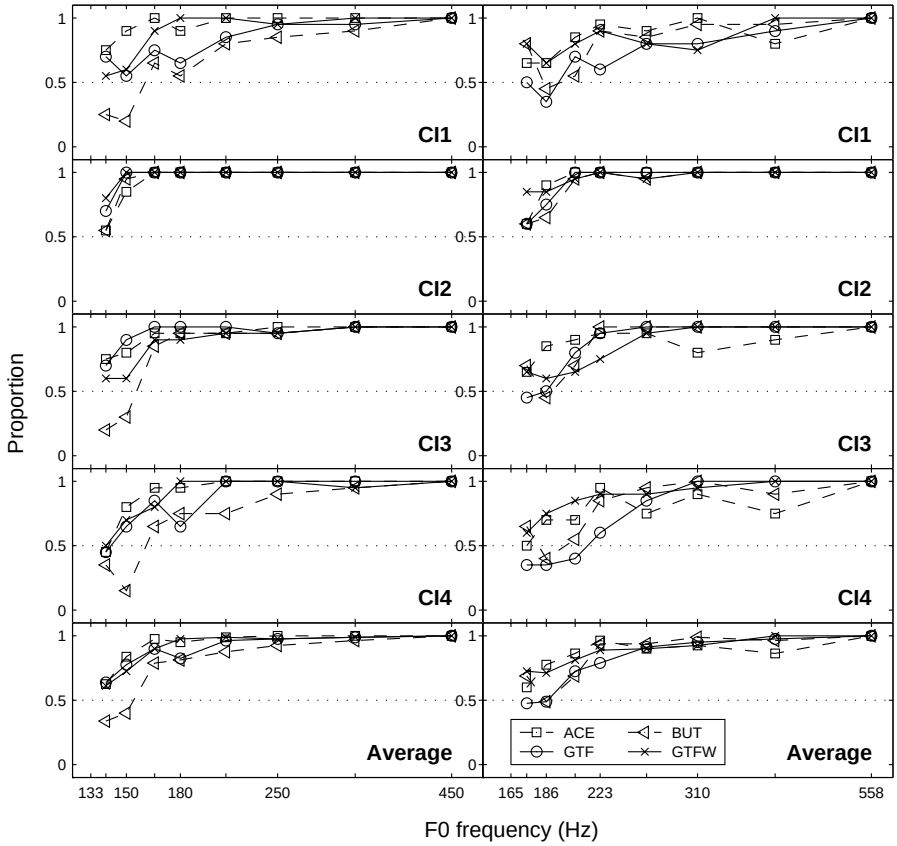


Figure 5.6: Results with both place and temporal pitch cues. The average proportion of correctly pitch ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations included up to 200 Hz. The fundamental frequencies of the signals are plotted on the horizontal axis on a log-scale. The left and right columns of panels depict the results for the condition where the reference F0 was 133 Hz and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

proportions depicted in Figure 5.5 and Figure 5.6 are based upon 20 presentations per data point.

A repeated measures analysis of variance (ANOVA) was performed on the proportions of correctly ranked trials using five factors. The independent variables were presence of temporal pitch cues (absent or present), reference F0 (133 or 165 Hz), filter bank (ACE, GTF, BUT, or GTFW), relative F0 difference (8 levels) and formant frequencies (five different F1). The main effects of presence of temporal pitch cues (p = 0.002) and relative F0 difference (p = 0.007) were significant. The other main factors were not significant: reference F0 (p = 0.678), filter bank (p =

0.251), and formant frequency (p = 0.122). The latter result allows averaging over the different formant frequencies in the presentation of the data in Figure 5.5 and Figure 5.6. Only the interaction term between filter bank and presence of temporal pitch cues was significant (p = 0.015). No other interaction terms were significant, but there were significant inter subject differences ($p < 0.001$).

Figure 5.5 and Figure 5.6 indicate that subjects were able to rank the stimuli based upon their fundamental frequency. They also show that as the relative F0 difference increases the difference is perceived more clearly. Also, comparing Figure 5.5 and Figure 5.6, the results indicate that providing temporal fluctuations in the envelope of the channels up to 200 Hz improves the ability of the subjects to discriminate F0 differences. This improvement is however dependent on the filter bank used, and the largest improvement is found for the ACE filter bank.

Because of the significant interaction between presence/absence of temporal pitch cues and the filter banks, the results are presented in two parts: with or without temporal pitch cues.

5.3.1 Without temporal pitch cues

A four-factorial repeated-measures analysis of variance was performed on the proportions of correctly pitch-ranked trials that were collected with stimuli without temporal pitch cues. The main effects of filter banks (p = 0.030) and relative F0 difference (p = 0.010) were significant. The effects of formant frequency and reference F0 were not significant, as well as all interaction terms. The between-subject differences were significant ($p < 0.001$).

Pairwise comparisons were performed using the least-significant-difference (LSD) on the filter bank design. The GTF and BUT filter bank produced significantly better performance than the ACE filter bank (p = 0.035 and p = 0.018, respectively for GTF and BUT) and the GTFW filter bank (p = 0.049 and p = 0.015, respectively for GTF and BUT), but there was no significant difference between BUT and GTF. The GTFW filter bank allowed for significantly better performance than the ACE filter bank (p = 0.046).

The results indicate that even without temporal pitch cues it was still possible for the subjects to rank the stimuli based on F0. However, it must be emphasized that this was only possible for relatively large differences in F0 of about one octave for the BUT and GTF filter bank and more than one octave for the GTFW filter bank. For the ACE filter bank, the subjects found it almost impossible to use place pitch cues to rank F0, even when stimuli differed more than 20 semitones.

5.3.2 Including temporal pitch cues

A four-factorial repeated measures ANOVA was performed on the proportions of correctly ranked trials where the stimuli now contained temporal pitch cues. The main factors reference F0 (p = 0.006) and relative F0 difference (p = 0.013) had a significant effect. No other main effects were significant. The effect of the filter bank was not significant (p = 0.126).

The discrimination was significantly poorer at the higher reference F0 tested. This result agrees with previous studies showing that the effectiveness of temporal pitch cues decrease with increasing overall rate [Shannon, 1983a].

Three of the four subjects (CI1, CI3 and CI4) did not reach perfect performance for the ACE filter bank and reference F0 of 165 Hz, until the relative F0 difference was more than one octave. Nevertheless, the performance was above chance level for relatively small F0 differences. This indicated that further increasing the F0 difference did not make the two processed stimuli perceptually more dissimilar. This was probably due to the poor resolution of the ACE filter bank in the F0 range causing F0 to be encoded solely by temporal pitch cues, whose effectiveness was limited at high F0s. This limitation might have been caused by the limited modulation depth of the envelopes at higher frequencies or by the reduced perceptual differences for temporal pitch cues at higher rates. The three IIR filter banks (GTF, BUT and GTFW) did not suffer from this shortcoming because for larger F0 differences the pitch discrimination was most probably also mediated by place pitch.

The difference in results between the condition without temporal pitch cues and the condition with both place and temporal pitch cues quantifies the influence of temporal pitch cues present in the presented stimuli. To asses this, the proportions of correctly ranked trials were averaged over the four subjects and converted into *d'* values for all conditions and all stimuli. To prevent infinite *d'* values for perfect responses, the perfect responses are adjusted by introducing an error of half a trial as suggested by MacMillan and Creelman [1991]. The amount of perceptual difference, introduced by the temporal pitch cues, is the difference of the *d'* values between the condition with temporal pitch cues ($d'_{combined}$) and the condition without temporal pitch cues (d'_{place}). The calculation of the discriminability based upon the temporal pitch cues is derived from Eq. (4.7) and given by:

$$d'_{temp} = sign(X)\sqrt{|X|} \quad , \tag{5.20}$$

with

$$X = sign(d'_{combined})(d'_{combined})^2 - sign(d'_{place})(d'_{place})^2 . \tag{5.21}$$

All data points where both conditions, with and without temporal pitch cues, resulted in perfect discrimination, were discarded because these points did not provide any information on the effect of adding temporal pitch cues. The additional perceptual effect based upon the temporal pitch cues is shown in Figure 5.7. Data points represent averages over formant frequency and reference F0. The temporal pitch cues become clearer as the F0 difference increases for all filter banks but the clarity of the temporal pitch cues differs between filter banks. The ACE and GTFW filter bank, having broad filters, allow for the largest increase in performance based upon temporal pitch cues. The GTF and BUT filter bank, which have narrower filters, exhibit less increase of the discrimination performance possibly because the baseline performance was higher.

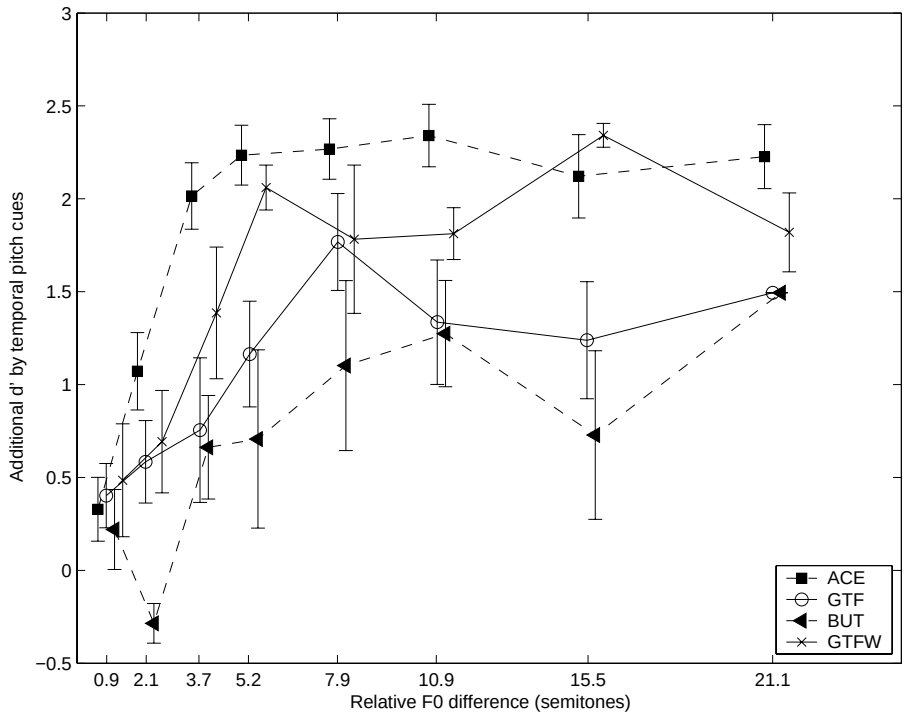


Figure 5.7: The additional d' caused by the inclusion of temporal pitch cues in the processed stimuli, for the four filter banks tested in this study. The additional d' values are calculated based upon the average response of the four CI subjects, and are derived from the difference in d' between the corresponding data points with and without temporal pitch cues (see text for details). The additional d' is averaged over the 5 different formant frequencies and over both reference F0 conditions. The error bars indicate the standard error of the mean. The data points are slightly shifted along the x-axis for clarity.

5.4 Discussion

5.4.1 Without temporal pitch cues

Previous research has shown that F0 discrimination can take place in the absence of temporal pitch cues in CI subjects [Geurts and Wouters, 2001; Geurts and Wouters, 2004]. This is confirmed by our results. Additionally, it is shown that this ability to discriminate F0 based on place pitch cues, is highly dependent on the filter bank used. The ACE filter bank did not allow the F0 information to be coded by place pitch cues. However, the BUT, GTF and GTFW filter banks, which have more resolution in the lower frequencies, did enable the subjects to perceive F0 differences based upon place pitch. The GTFW has slightly less resolution compared to GTF and BUT due to the large filter overlap, and consequently the F0 discrimination using this filter bank is somewhat lower compared to the latter two

filter banks. The actual shape of the filter bank, being more rectangular with steep slopes and less overlap (BUT) or more triangular with shallower slopes and more overlap (GTF), did not significantly change the performance. This contrasts with the results of Geurts and Wouters [2004] who found an improved performance for a triangular-like filter bank. However, given the present results, the improved performance of the triangular filter bank in that study was probably facilitated by the different cut-off frequencies used in the triangular filter bank, leading to more resolution in the lower frequency register. For subjects that have poor temporal pitch sensitivity but good place pitch sensitivity the filter banks with high resolution in the F0 register may enable them to better discriminate F0.

Some of the psychometric curves of Figure 5.5 are non-monotonic and contain pitch reversals, some of which are consistent over the various subjects. In addition, when pooling the data for the different subjects and plotting the separate psychometric curves for the different F1 values, some significant pitch reversals can be observed, e.g. the psychometric curve for F1 = 500 Hz in Figure 5.8A. An analysis of all conditions that resulted in significant pitch reversals after pooling the data over the different subjects reveals a consistent pattern of the average channel amplitude before compression. This pattern is most easily interpreted by means of an example. Figure 5.8B shows the average amplitude of the 22 channels for a stimulus with F1 = 500Hz, processed with the GTF filter bank. The thick line represents the normalized average amplitude for the reference stimulus having an F0 of 133 Hz. The two thin lines represent the normalized average amplitude for two comparison stimuli, F0 = 150 Hz (solid line) and F0 = 250 Hz (dashed line). The former stimulus resulted in a pitch reversal and the latter stimulus resulted in a correct pitch order (the arrows in Figure 5.8A). When comparing F0s 133 and 150 Hz , the position of the fundamental frequency moves from the first to the second channel (towards the base), but the peak caused by the harmonics under the formant frequency shifts downward from channel 7 to channel 6, towards the apex. So the pitch percept was not determined by the place of the fundamental frequency, but was largely determined by the place of the largest excitation along the electrode array. In the case where the channels with the largest amplitude coincide, as in Figure 5.8B for F0 133 Hz (thick solid line) and F0 250 Hz (thin dashed line), the pitch difference was most probably mediated by the excitation shift along the electrode array of the smaller peaks. Summarizing the two previous effects, we hypothesize that the pitch of the different sounds was determined by a weighted position of the peak of the excitation pattern. The simplest form of a weighted peak position was the center of gravity or the centroid C^{elec} , and can be expressed as:

$$C^{elec} = \frac{\sum_{n=1}^{22} n.\overline{A_n}}{\sum_{n=1}^{22} \overline{A_n}} \tag{5.22}$$

In Eq. (5.22) n indicates the channel number (ranked from apex to base) and $\overline{A_n}$ is the average amplitude on channel n :

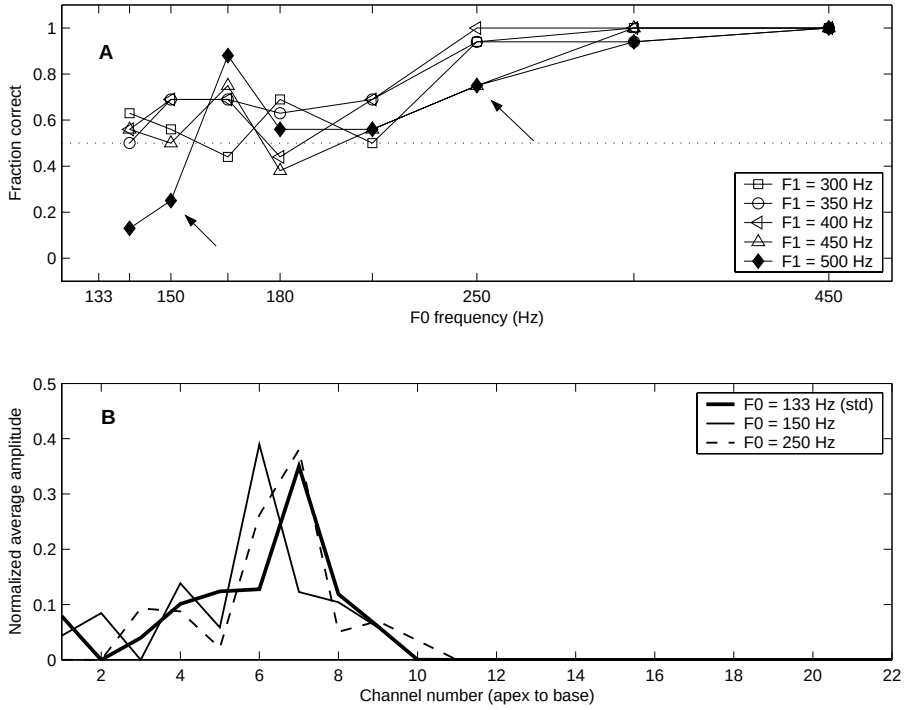


Figure 5.8: Top panel: The proportion of correctly pitch ranked stimuli pooled over all subjects for the condition with reference F0 133 Hz and stimuli processed with the GTF filter bank and all temporal modulations above 10 Hz removed. The different lines in the figure depict the results for the various formant frequencies used. The fundamental frequencies of the signals are plotted on the horizontal axis on a log-scale. A significant pitch reversal is noticed for the case where the formant frequency is 500 Hz (filled symbols) for the two lowest tested F0 frequencies (140 and 150 Hz).

Bottom panel: The average channel amplitudes for the single formant stimuli with formant frequency 500 Hz and fundamental frequencies 133, 150 and 250 Hz when processed with the GTF filter bank and with all temporal modulations above 10 Hz removed. The thick line represents the average amplitude of the reference stimulus with F0 133Hz, F1 500 Hz and processed with the GTF filter bank (i.e. the reference stimulus of the line with the filled symbols in the upper panel). The thin solid line represents the average channel amplitude of the comparison stimulus mentioned in the upper panel and causing a pitch reversal (F0 = 150 Hz). The peak of the average amplitude has shifted towards the apex with respect to the average amplitude of the reference stimulus (thick line), causing the comparison stimulus to be perceived at a lower pitch. The thin dotted line represents the average amplitude of a comparison stimulus correctly ranked in pitch (F0 = 250 Hz). Although the peak of the average amplitudes of the reference and comparison stimulus are at the same location, most probably the correct ranking is due to the shift of the fainter channels towards the base.

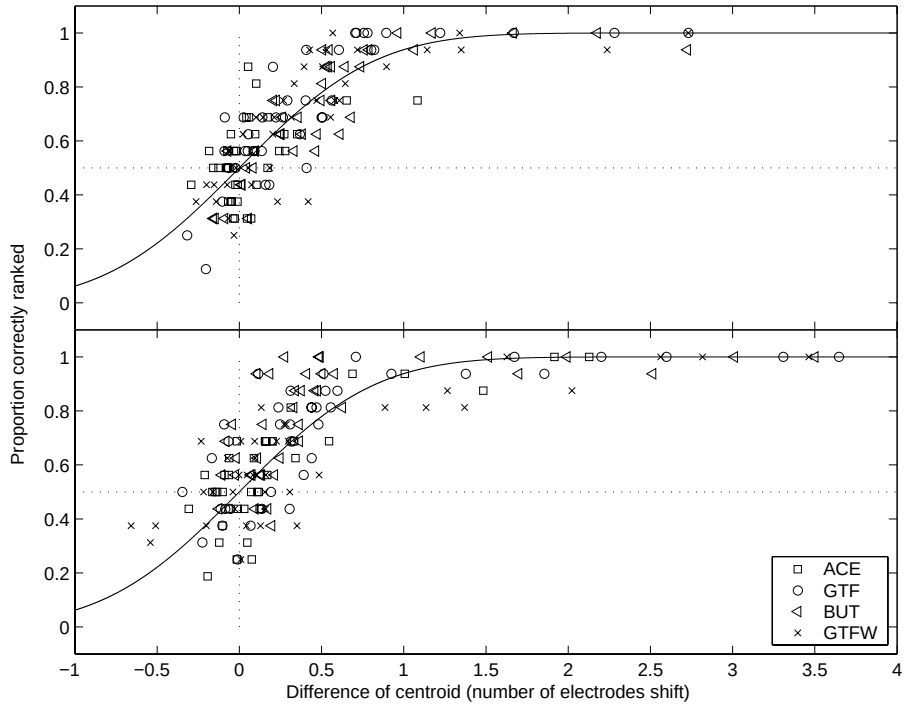


Figure 5.9: The proportion of correctly ranked stimuli pooled over all subjects plotted as a function of the difference in centroid. Only data of stimuli that allowed no temporal pitch cues was included in this picture. For the sake of clarity the data for the two reference F0 is depicted in separate panels (F0 133 Hz in the upper panel and F0 165 in the lower panel). A normal cumulative distribution function was fitted to the pooled data and is represented by the solid line on both panels.

$$\overline{A_n} = \frac{1}{T_{stimulus}} \sum_{t}^{T_{stimulus}} A_n(t), \tag{5.23}$$

where $T_{stimulus}$ is the total duration of the stimulus.

The centroid is computed based upon the average amplitude of the envelopes (the compression accommodating for loudness growth is not applied). According to the proposed model, the discriminability of two sounds (based upon place pitch cues) was determined by the distance between the centroids of the excitation patterns of the two sounds. The centroid has the additional benefit of being independent of overall intensity. Figure 5.9 shows a scatter plot of the model prediction of all stimuli and the respective average proportion of correctly ranked trials, for F0s of 133 Hz (top panel) and 165 Hz (bottom panel). The solid line in Figure 5.9 represents the normal cumulative distribution fit to the pooled data over all conditions and subjects. This function was fitted to the average proportion using a Gauss-Newton nonlinear least-squares fitting method.

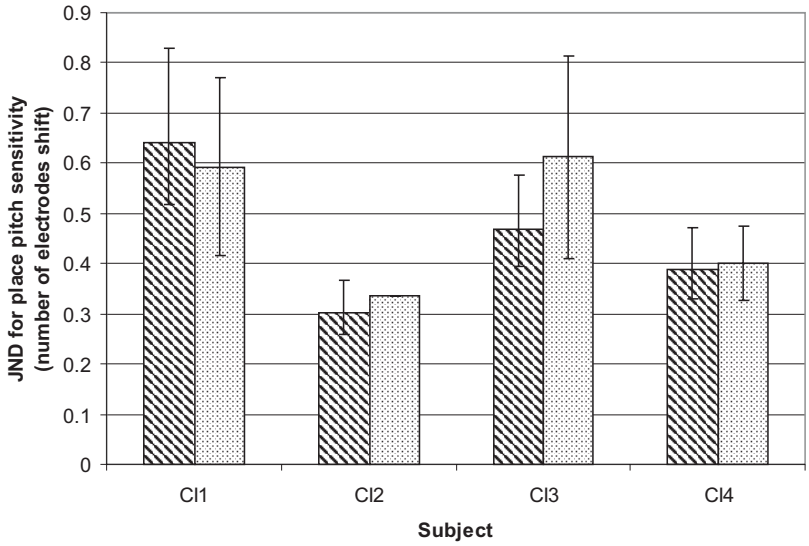


Figure 5.10: Estimated just noticeable differences (JNDs) for the four subjects. The dashed bars present the JNDs from the centroid differences calculated from the F0-discrimination data. The JNDs estimated from site of stimulation discrimination of the multi-channel stimuli in the same patients (see chapter 3) are depicted by means of the dotted bars. The error bars indicate the 95% confidence interval on the JND estimated given by the fitting routine for the centroid JNDs. The error bars on the place pitch sensitivity of the multi-electrode stimuli indicate ± 1 standard deviation.

Similarly, a normal cumulative distribution function was fitted to the scatter plots for every subject separately. The slope of this fitted normal cumulative distribution function was taken as a performance measure of place pitch sensitivity. This fitting procedure is similar to a linear regression with the data points converted into d' values. JNDs were derived from the fitted curves as the minimal distance expressed in number of electrodes to obtain 75% correct pitch rankings.

Figure 5.10 depicts the JND estimates that were obtained from these fitted curves for every subject. The place pitch sensitivity of the subjects was also measured using an electrode ranking experiment, as described in Chapter 3. These JND estimates are also depicted in Figure 5.10. In the study of Chapter 3, the subjects were presented two stimuli that differed in site of stimulation along the electrode array, and had to indicate the highest one. The JNDs shown in Figure 5.10 are the mean of the JNDs for conditions with different numbers of active channels. The results of chapter 3 did not demonstrate a significant effect of the number of active channels of the stimuli on place pitch sensitivity.

The JNDs estimated from the F0 discrimination data and the JNDs estimated from the place pitch measurements are nearly equal for all subjects (within error bars).

However, for subject CI3, the place pitch sensitivity is somewhat worse when compared to the centroid sensitivity. This can be due to the fact that we arbitrarily choose two electrode locations to estimate the place pitch sensitivity. At other locations the place pitch sensitivity for subject CI3 might be better.

Previous support for the centroid model has been presented by McDermott and McKay [1994] for the case of 2 stimulation channels in cochlear implants. They investigated the pitch perceived when 2 bipolar channels were stimulated simultaneously and interleaved using constant current levels. The pitch of the two-channel stimuli varied monotonically between the pitches of the stimuli on the two single electrodes, as the relative current amplitudes were changed in orderly fashion. They interpreted their results by stating that the centroid of the excitation pattern shifted from one electrode location to the other as relative currents varied and that this centroid shift caused the varying pitch. Their results also indicated that the pitch of two interleaved stimulated channels was perceived as one-dimensional, at least up to separations of 3 mm between stimulating electrodes. McKay *et al.* [1996] later confirmed and extended this result using dissimilarity matrices and multidimensional scaling analysis. They concluded that there exist two basic perceptual dimensions with two-channel stimulation. The first percept was associated to the middle of the two active channels, corresponding to the centroid of the excitation pattern. The second percept was associated to the separation of the two active channels. However, McKay *et al.* also argued their experiment did not rule out that the dimensions could also be interpreted as simply being associated with the location of the two channels. The results of McKay *et al.* [1996] were consistent for separations up to 12 mm, the largest separation tested, between active channels. Due to the nature of the acoustic stimuli used in the present study and the maxima selection in the sound processing stage, the pulse sequences only stimulated channels that were at most 10 electrodes apart, resulting in a maximal separation of 7.5 mm between active channels.

Cohen *et al.* [1996a] examined the direct relationship between the forward masking pattern and the numerical pitch estimates that single channel stimuli exert. They found a clear qualitative correspondence between the centroid of the forward masking pattern and the numerical pitch estimates, for all stimulation modes. They were also able to predict the reversed (compared to normal tonotopy) pitch ranking of some pairs of electrodes based upon the centroid of the forward masking pattern for several subjects.

In normal-hearing subjects, models based upon the centroid of the spectrum [Stover and Feth, 1983; Anantharaman et al., 1993] have been used to explain the results of pitch ranking and discrimination experiments with noise bands [Dai et al., 1996; Versfeld, 1997]. These models accurately described the discriminability of narrow band sounds, but the models had to be extended for the discrimination of sounds with larger bandwidths.

It is, however, questionable whether the place pitch is truly a pitch percept or an aspect of the sounds' timbre. According to the ANSI definition of pitch [American National Standards Institute, 1994], place pitch is pitch as subjects are able to rank stimuli from high to low. Nevertheless, there is a striking resemblance between the

proposed model for place pitch and the model for sharpness in normal-hearing subjects [Zwicker and Fastl, 1999]. Both models calculate the centroid of the excitation pattern.

The results of the present study and the studies of Geurts and Wouters [2001; 2004] have shown that F0 discrimination in cochlear implant recipients is possible solely based upon place pitch cues. These studies used stationary synthetic signals while running speech is non-stationary and formant frequencies change over time. Green *et al.* [2002] measured the discrimination of direction of sweeps of fundamental frequency using noise-band vocoders in normal-hearing subjects. They showed that for stationary signals the discrimination was possible based solely on place pitch cues but the discrimination became almost impossible when the signals had non-stationary formant frequencies. Moreover, the model presented above to account for the F0-discrimination results based upon place pitch cues, indicates a strong dependence of the place pitch cues on the overall spectral shape of the stimuli. This might limit the usefulness of place pitch cues to encode F0 in cochlear implants during running speech.

On the other hand, although formant frequencies and F0 frequencies originate from physically independent processes, formant frequencies and F0 frequencies are often correlated. E.g. both the formant frequencies and the F0 frequencies of a female speaker are on average higher than those of a male speaker. The combined effect of higher formant frequencies and higher fundamental frequencies will shift the centroid of the excitation pattern toward the base. This shift might be detectable for cochlear implant subjects even during running speech. This might help the subject listening to running speech to discriminate pitch differences.

Moreover, the possible problem for F0-discrimination based upon place pitch cues coming from the interaction between overall spectral shape and F0 cues is less likely to exist for some musical sounds, such as woodwinds, because these musical notes have stationary spectral shapes throughout their duration [Grey, 1977]. Secondly, the dimensions and form of the instrument, defining the resonance frequencies, are fixed so the spectral shape is fixed across different notes. Consequently, the differences in place pitch between two different notes are related to F0 differences.

5.4.2 Including temporal pitch cues

Comparing Figure 5.5 and Figure 5.6, the improvement in performance from adding the temporal pitch cues was largest for the ACE and GTFW filter banks. However, the range for improvement was a lot larger for the ACE filter bank than for the other filter banks because the latter already enabled some F0 discrimination without temporal pitch cues.

The limited improvement obtained by adding temporal pitch for the GTF and BUT filter bank might have been due to the fact that higher harmonics of the test stimuli are also resolved in those filter banks. Resolved harmonics prevented the output of filters from beating at the fundamental frequency, and the only fluctuation seen at the output was at the frequency of the respective harmonic. The filter outputs with the largest amplitude were filters having center frequencies close to the formant frequencies (between 300 and 500 Hz). In this range the BUT and GTF filter banks

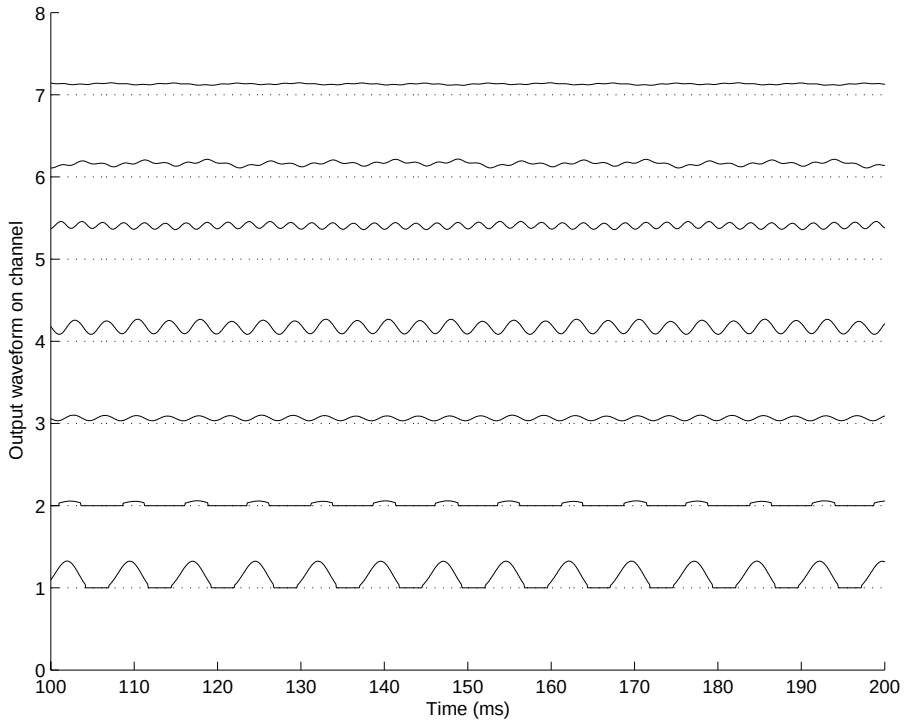


Figure 5.11: Time fragment of the amplitudes of channels 1 to 7 across time for the SFS stimulus with F0 = 133 Hz and F1 = 400 Hz, processed with the BUT filter bank. Channels 1 and 2 only contain the first harmonic, and consequently the outputs beat at the fundamental frequency. Channels 3 and 4 only contain the second harmonic, leading to a modulation at the second harmonic frequency. Channel 5 contains only the third harmonic, and consequently the output beats at the third harmonic frequency. The beating at higher harmonics makes the temporal pitch cues less clear.

resolved the individual harmonics, so the largest output of the filters fluctuated at multiples of the fundamental frequency. As an example, the modulations of channels 1 to 7 of one of the sounds processed with the BUT filter bank and after envelope detection are depicted in Figure 5.11. The most intense channels in this example do not beat at F0 frequency but at some higher harmonic frequency. Although this higher frequency fluctuation was partly attenuated by the envelope detection low-pass filter, it is still clearly present in the channel output. Only the output of the lower channel, containing the fundamental frequency, and possibly higher channels, containing unresolved harmonics, provide the temporal pitch cues related to F0. The channels containing resolved higher harmonics are modulated at multiples of F0 and consequently elicit higher pitches because higher modulation frequencies lead to higher pitches [McKay et al., 1994].

The ACE filter bank did not suffer from this shortcoming because all the bins in the FFT filter bank are approximately 250 Hz wide so all harmonics were unresolved (at least for the reference stimuli), causing all filter outputs to beat at the fundamental

frequency. Similarly, the GTFW filter bank was designed to only resolve the fundamental frequency and with all overtones unresolved. This also made the output of all channels beat at F0 for the GTFW filter bank.

In order to assess the effect of the modulation depth of the beatings at F0 in the different channels, the average modulation depths were computed of all stimuli. We also assess the effect of the phase differences of the beating across channels. These inter-channel phase differences of the beating at F0 are caused by the phase differences between the different harmonics of the original sound and by the phase response across frequency of the filters in the filter bank. The average modulation depth over all channels is computed in two different ways to assess the effect of the phase differences. In the first way, all phase information of the channels was discarded ("without phase"). In the second way the phase information was included and contributed to the average as out-of-phase channels could interfere destructively ("with phase").

As a first step, the average spectrum $\Phi_n(q)$ of the modulations in all channels was calculated.

$$\Phi_n(q) = \sum_{t=0}^{T_{stimulus}} A_n(t) e^{-j\frac{2\pi}{T_{stimulus}}tq} \quad , n = 1...22 \; . \tag{5.24}$$

For the 'without phase' approach, the spectrum of all channels was calculated separately and the average of the magnitude of the spectrum over the channels was taken, discarding the phase components of the modulation:

$$\left|\overline{\Phi}_{without\ phase}(q)\right| = \frac{1}{N}\sum_{n=1}^{N}\left|\Phi_n(q)\right| . \tag{5.25}$$

In the 'with phase' approach, the time signal was averaged over all channels and then the magnitude of the spectrum of this average time signal was calculated.

$$\left|\overline{\Phi}_{with\ phase}(q)\right| = \frac{1}{N}\left|\sum_{n=1}^{N}\Phi_n(q)\right| . \tag{5.26}$$

In the second step, the modulation depth was calculated for both approaches from the respective average spectrum, using:

$$ModulationDepth = \frac{2\left|\overline{\Phi}(F0)\right|}{\left|\overline{\Phi}(F0)\right| + \left|\overline{\Phi}(0)\right|} \tag{5.27}$$

where $\left|\overline{\Phi}(F0)\right|$ was the magnitude of the spectrum at the fundamental frequency and $\left|\overline{\Phi}(0)\right|$ was the magnitude of the DC component of the spectrum.

As a third step, the obtained modulation depths were correlated with the amount of temporal pitch cues (d'_{temp}) for all processed reference stimuli and all test conditions. The correlation was done only with the modulation depth of the

reference stimuli and not with the modulation depth of the higher F0 comparison stimuli, because of the test procedure used in the present experiment. The subjects were asked to indicate the highest sounding stimulus and this could be the one with a higher modulation rate or with no modulation at all. This is because for modulated stimuli the perceived pitch is related to the modulation frequency but steady state (or poorly modulated) stimuli elicit high pitches related to the high overall pulse rate [McKay and Carlyon, 1999]. So, clear pitch differences may have been perceived when the comparison contained no modulations at all.

The analysis revealed that the d'_{temp} values are significantly correlated ($p < 0.0001$ for both) with the modulation depth of the reference stimulus of the trials for both computational methods ($r = 0.4216$; and $r = 0.3090$ for 'with phase' and 'without phase' respectively). Moreover, although both methods of calculating the modulation depth are significantly correlated ($r = 0.5910$) with each other, the correlation with benefit from inclusion of temporal pitch cues was 36% higher when phase information was included. Using a test comparing the two sets of predictors [Tabachnick and Fidell, 1996], this lead to a significantly higher correlation for the 'with phase' approach ($Z = 2.42$; $p = 0.0077$). However, both correlations are rather weak and the results of this analysis should be interpreted with some caution.

The poorest performing filter bank for temporal pitch cues (BUT) produced the lowest modulation depths averaged over all reference stimuli. The best performing filter bank for temporal pitch cues (ACE) produced the highest average modulation depth for the reference stimuli, when modulation depths are calculated taking into account the relative phase differences between the channels. When the phase information is discarded, GTF produced the lowest modulation depths and GTFW the highest modulation depths. So, the 'with phase' modulation depth present in the processed stimuli can be a predictor of the amount of temporal pitch cues that can be used by the cochlear implantee.

This may mean that out-of-phase modulations on different channels probably allow for less clear pitch percepts than equally large in-phase modulations, at least when the channels are within a relatively narrow region, as is the case in this study. It has been shown before that modulation patterns on two nearby electrodes are integrated, while more widely separated modulation patterns are perceived separately [McKay and McDermott, 1996].

5.5 Conclusions

In this chapter we examined properties of filter banks affecting F0 discrimination in cochlear implants. Temporal and place pitch cues to F0 discrimination were studied separately and for both cues the mechanisms relating the pitch percepts to physical parameters were investigated.

The results of our experiments demonstrate that both place and temporal pitch cues allow the discrimination of the fundamental frequency, as also evidenced by previous research. However, F0 discrimination based upon solely place pitch cues is weak and only possible for differences exceeding one octave for the best performing filter banks. Moreover, the ACE filter bank, as used by most of the Nucleus cochlear

implant recipients in their daily life, did not allow for F0 discrimination based upon place pitch cues. Our experiments indicate that our new filter banks having the highest frequency resolution in the F0 register gave the best results using place pitch cues. The newly proposed filter banks are likely to improve F0 discrimination, relative to the currently clinically available filter bank, in subjects having good place sensitivity and poor temporal sensitivity, at least for spectrally stationary signals. A previous study has shown that for some cochlear implantees the place pitch cues are very important for F0 discrimination [Geurts and Wouters, 2004].

When temporal pitch cues were added to the stimuli, performance increased significantly and JNDs for F0 discrimination ranged from 6% up to 60% depending on subject, processing condition and reference F0. There were no overall significant differences between the performances of the filter banks when both temporal and place pitch cues were presented to the subjects. The performance of cochlear implant subjects is still very poor compared to the performance of normal-hearing subjects where JNDs for F0 are typically less than 1%. The largest temporal pitch cues for F0 discrimination were possible with the filter banks that did not resolve the harmonics. In these filter banks all output channels beat at the fundamental frequency. When the beatings of the different channels were in phase, the temporal pitch cues seemed to be clearer compared to the case where beatings of neighboring channels are out-of-phase.

A modeling study of the results with solely place pitch cues indicated that place pitch is related to the centroid of the excitation pattern along the cochlea. As place pitch cues enable subjects to rank stimuli from low to high it can be considered a pitch percept. However, taking into account the present model, place pitch is more related to the sharpness or brightness of timbre than to the repetition rate of the processed sound in cochlear implants.

The proposed model indicates that the place-pitch-mediated F0 cues and the overall spectral shape are closely related. In other words, there might be interference in the perception between F0 and formant frequencies when both are coded onto place of stimulation. This might possibly lead to a reduction of the usefulness of the place-pitch-mediated F0 cues when formants vary, as in running speech, and might possibly also lead to reduced speech perception when F0 strongly varies. Although the spectral shape or brightness of timbre also interferes with pitch perception in normal-hearing subjects [Ohgushi, 1978], the effect in cochlear implants appears to be larger at least when only place pitch cues are present.

With the filter banks studied, optimal transmission of both temporal and place pitch cues was difficult for one and the same filter bank.

Chapter 6

Acoustic models for pitch perception by cochlear implant subjects

J. Laneau, M. Moonen, and J. Wouters. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *Journal of the Acoustical Society of America,* In review.

Pitch-perception research is not feasible without experiments with subjects. However, sometimes it might be better to use an (acoustic) model than to measure directly from CI subjects. An acoustic model is as a mechanism that generates acoustic stimuli eliciting similar perception of sounds in NH listeners as in CI subjects using their device. In other words, an acoustic model presents a simulation of the sound sensation cochlear implantees experience to NH listeners. Most often acoustic models are built using signal processing.

There are a number of benefits to the approach of simulating a CI in a normal-hearing person. For instance, in an acoustic model a number of parameters such as insertion depth of the electrode array can be varied where this is impossible in CI

subjects. An acoustic model also gives the (normal-hearing) researcher an idea of the sound CI-subjects actually hear. A more detailed enumeration of the benefits of acoustic models is found in section 6.1 .

Acoustic models of CI have been designed and used experimentally in previous studies. In a number of experiments, *noise-band vocoders* have been shown to provide adequate acoustic models for speech understanding in CI. However, this chapter assesses whether and under what limitations these acoustic models can be used for pitch perception research in CI.

Section 6.2 describes the signal processing in the noise-band vocoder used in this chapter as acoustic model for pitch research in CI. In sections 6.3 and 6.4, the experiments examine the effect of center frequency and spectral slopes of the noise carriers of the vocoder on simulated place and temporal pitch sensitivity, respectively. In section 6.5, the data from CI subjects on fundamental frequency discrimination from chapter 5 are directly compared to data of normal-hearing subjects presented with the same stimuli processed with two different noise-band vocoders. The results show that place pitch sensitivity drops when the spectral slopes become shallower, and that place pitch cues for multi-channel stimuli can adequately be mimicked when the discriminability of adjacent channels is adjusted by varying the spectral slopes to match that of CI subjects. The spectral smearing caused by the shallow slopes of the noise-band carriers affects pitch perception more than speech perception. The results also indicate that temporal pitch sensitivity is limited for noise-band carriers with low center frequencies.

6.1 Introduction

Previous studies have shown that noise-band vocoders can be used as an acoustic model for speech perception in cochlear implants [Blamey et al., 1984a; Dorman and Loizou, 1997; Dorman and Loizou, 1998; Friesen et al., 2001; Nelson et al., 2003b]. These studies have indicated that at least good qualitative agreement exists between the model results and the results obtained from CI subjects as along as the number of channels is below eight. More recently, several studies have used noise-band vocoders to study pitch perception in CI [Green et al., 2002] or pitch related tasks [Fu et al., 1998; Faulkner et al., 2000; Xu et al., 2002; Luo and Fu, 2004; Kong et al., 2004; Fu et al., 2004]. However, it is generally unknown whether noise-band vocoders provide a valid acoustic model to study pitch perception in CI. For it has been shown that speech perception relies on different mechanisms than pitch perception [Smith et al., 2002]. Smith *et al.* showed that while speech perception is mostly based upon envelope information in broad frequency bands, pitch perception is based upon the fine structure within every frequency band. In this chapter, pitch discrimination results from CI subjects are compared to results from normal-hearing (NH) subjects listening to noise-band vocoder processed signals measured under the same conditions, to assess the accuracy of the acoustic model.

There are several benefits of using acoustic models in parallel to performing experiments directly with cochlear implant recipients. Firstly, the results obtained from cochlear implant patients are generally more prone to inter- and intra-subject variability [Friesen et al., 2001]. Possible reasons for this higher variability might be

the different etiologies of the patients, differences in duration of auditory deprivation leading to various amounts of degeneration of the auditory functions, and the high variability in peripheral processes, such as neural excitation profiles, between subjects and even within subjects. A second benefit of using acoustic models is the fact that in acoustic models most of the parameters can be changed independently, whereas in cochlear implant patients some these factors are fixed, such as width of excitation pattern or implant insertion depth. Thirdly, comparing the results of normal-hearing subjects using acoustic models on the one hand and cochlear implant patients on the other hand reveals some details of the basic functioning of the auditory system. Basic mechanisms of the auditory system can be more solidly founded when the results in electrical and acoustical hearing reveal the same trends [Carlyon et al., 2002; van Wieringen et al., 2003]. Fourthly, comparing results obtained through acoustic models and results obtained directly from CI recipients also allows indicating the cause of limitations in CI recipients' performance. For instance, when the normal-hearing persons have the same limitations as CI subjects, the limitations are caused by a part of the CI system (or possibly by a feature of the resynthesis in the acoustic model). Finally, there is a practical benefit to acoustic models because the pool of normal-hearing subjects is vastly more extended than the pool of CI subjects, and research interfaces to stimulate acoustically are more easily available than research interfaces to stimulate implants directly.

Previously several acoustic models have been developed and used to study pitch perception in cochlear implants through normal-hearing subjects. Most of these acoustic models can be categorized in three groups:

1) Noise-band vocoders

Firstly, noise-band vocoders are most often used [Fu et al., 1998; Faulkner et al., 2000; Green et al., 2002; Xu et al., 2002; Kong et al., 2004] because temporal and place pitch cues appear to be accurately modeled. Modulation of noise carriers elicits weak purely temporal pitch percepts that correspond to the modulation frequency [Burns and Viemeister, 1976] like temporal pitch in CI [McKay et al., 1994]. Moreover, the difference limens in rate discrimination for CI resemble the modulation rate discrimination of noise carriers as a function of frequency [Shannon, 1983a; Blamey et al., 1984b]. Furthermore, the place pitch cues can be modeled by changing the center frequency of the noise band because the pitch of a steady-state noise band corresponds approximately to its center frequency [Stover and Feth, 1983; Dai et al., 1996; Zwicker and Fastl, 1999].

2) Sinusoidal vocoders

Whereas noise-band vocoder signals sound very stochastic, CI subjects never define the perceived sound as noisy but rather qualify the perceived sound of single channels as beeps. To accommodate for this fact a second acoustic model was developed that uses pure tones as carriers for the different channels in the resynthesis of the vocoder [Dorman et al., 1997]. More recently, such a vocoder was used to study the pitch perception through discrimination of lexical tones for two different sound processing schemes [Lan et al., 2004]. The sinusoidal vocoder might be less accurate in modeling pitch perception in CI because

adjacent channels are extremely well discriminable in the model while in general CI subjects can have problems discriminating adjacent channels [Nelson et al., 1995]. Furthermore, amplitude modulation of a pure-tone carrier generates sidebands which can be detected when the sidebands are spaced wide enough to be resolved by the peripheral filters of the normal ear [Kohlrausch et al., 2000], while this is impossible in CI subjects.

3) Filtered-harmonics vocoder

A third acoustic model was proposed to study the purely temporal pitch percepts and uses harmonic complexes that are band-pass filtered in relatively high frequency regions so that all harmonics are unresolved [McKay and Carlyon, 1999; Carlyon et al., 2002; van Wieringen et al., 2003; Deeks and Carlyon, 2004]. Using this scheme place and temporal pitch cues can be adjusted independently by varying the center frequency of the band-pass filter and the fundamental frequency of the harmonic complex respectively. However, the model is restricted to lower pulse rates because for higher pulse rates the harmonics become resolved.

In this chapter we chose to use and design a noise-band vocoder as an acoustic model for pitch research in CI because sinusoidal vocoders most probably provide too clear place pitch cues relative to place pitch cues in CI subjects, and because unresolved pulse trains are only feasible for low-pulse rates (below about 200 pps) while modern CI systems typically stimulate at much higher rates (above 900 pulses per second).

This chapter examines the factors that affect the use of a noise-band vocoder as an acoustic model of pitch perception in CI. In section 6.2 we describe the signal processing in the noise-band vocoder used in this chapter. The two first experiments (1A and 1B, in sections 6.3 and 6.4, respectively) focus on the resynthesis stage of the vocoder and assess the effects of center frequency and spectral slopes of the noise-band carriers on the simulated temporal and place pitch cues. In experiment 2 (section 6.5) F0-discrimination results for CI subjects from chapter 5 are compared to F0-discrimination results for normal-hearing subjects for two different vocoders. Experiment 2 thus includes the influence on F0-discrimination of variations to both the analysis and resynthesis stage of the vocoder.

6.2 Sound processing scheme

6.2.1 Analysis-resynthesis vocoder

The processing of the noise-band vocoders can be divided in an analysis step and a resynthesis step. A block diagram depicting the different processing steps is shown in Figure 6.1. The analysis part is based upon the first processing stages of the processing in CI and basically splits up the input sound into a number of channels through a filter bank. The filter bank, envelope detection and maxima selection are identical to the sound processing of the modeled CI system. The vocoder does not contain a compression stage in the analysis part, in contrast to CI sound processing. The compression stage in CI systems is designed to map the relatively wide dynamic

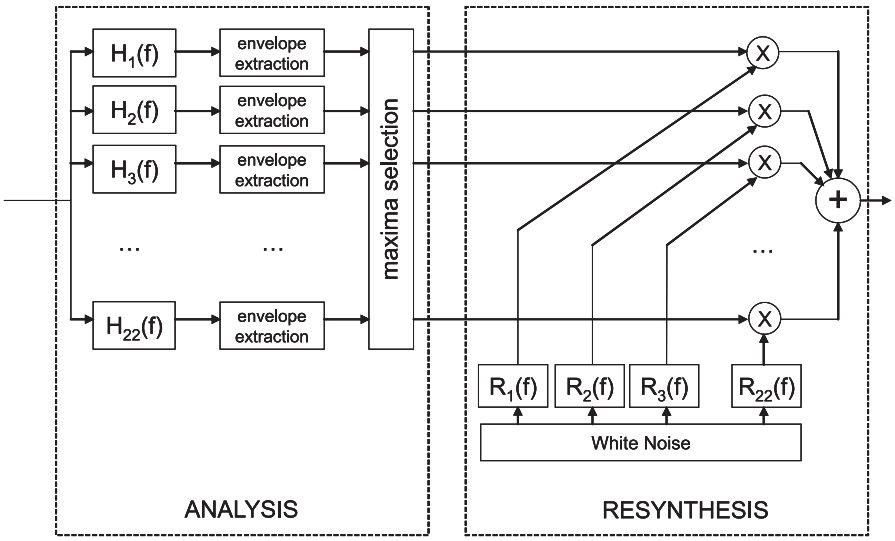


Figure 6.1: Block diagram of the processing scheme of the noise-band vocoders.

range of speech sounds onto the narrow dynamic range of electrically evoked hearing and to produce a normal or nearly normal growth of loudness with increases of signal level [Zeng and Shannon, 1994]. However, we chose not to include the compression stage into the vocoder because the dynamic range of normal-hearing persons is relatively wide and they, by definition, exhibit normal growth of loudness. Any compression or expansion of the envelopes would alter this normal growth of loudness and this would create a difference between sensation with a vocoder and sensation with a CI because we assume that in CI subjects (with the inclusion of compression) loudness growth is approximately normal. All other studies using vocoders to assess hearing with CI also exclude a compression/expansion stage unless they are specifically assessing the effect of envelope compression on sensation.

In the resynthesis stage of the noise-band vocoder, the output channels of the analysis stage are integrated again into one single audio-channel. This is generally done by modulating each channel with noise bands and summing up all channels. The design of the resynthesis stage is based upon the psychophysics and physiology of both electric hearing and acoustic hearing. Psychophysical studies have shown that temporal resolution of envelope modulation is similar for noise bands in normal-hearing persons and for pulse trains in CI recipients [Shannon, 1983a; Blamey et al., 1984b]. Increasing the band width of the noise bands allows simulating the channel interaction between adjacent channels [Shannon, 1983b] and the spectral smearing found in CI recipients caused by the current spread found with electrical stimulation in the cochlea [Shannon et al., 1998; Fu and Nogaki, 2005].

This design has been used in numerous other studies [for instance Shannon et al., 1995; Smith et al., 2002]. The ability to replicate results for speech perception obtained with CI subjects through simulations has been verified in a number of studies [Dorman and Loizou, 1997; Dorman and Loizou, 1998; Loizou et al., 2000;

Friesen et al., 2001]. The results of these studies show at least a good qualitative correspondence between CI results and simulation results when the channel number is below eight. This demonstrates the feasibility of a simulation of CI sensation for speech perception. This does not mean at all that the normal-hearing persons using the simulations experience the same sound as CI recipients do. However, the simulations provide the same (limited) speech intelligibility as in CI recipients.

6.2.2 CISIM resynthesis filter

In most noise-band vocoders the frequency response of the resynthesis filters is matched to the frequency response of the analysis filters (indicated by the 'standard vocoder' in section 6.5). In this way the spectral envelope of the processed sound is preserved and speech understanding is least affected by the processing [Shannon et al., 1998; Fu and Shannon, 1999; Baskent and Shannon, 2003]. However, in a cochlear implant the neural population excited by a particular stimulation channel (or electrode) does not necessarily corresponds to the neural population excited in the normal ear by the frequencies passing through this channel. For instance, the deepest inserted electrode of the electrode array is typically located about 10 mm from the apex of the cochlea. In a normal ear this location is stimulated by frequencies of approximately 500 Hz but in a cochlear implant system the deepest inserted electrode is typically stimulated for frequencies around 250 Hz. Also, the analysis filters in the CI sound processor typically have relatively steep slopes and the adjacent filters do not overlap significantly [Wilson et al., 1991]. However, in a cochlear implant the different stimulation channel overlap significantly in their excited neural populations.

In order to take into account the difference between analyzed frequencies and locations of excited neural populations, we propose to design a resynthesis filter that takes into account the location of the electrode inside the cochlea and the current spread obtained in the electrically stimulated cochlea: the CISIM filter (CI-SIMulation filter). This filter is loosely based upon the physiology of electrical stimulation through a cochlear implant. The center frequency of the filter is set to the frequency corresponding to the location of the simulated electrode. The shape of the frequency response of the filter is designed to mimic the exponential decay of current density along the basilar membrane [Black and Clark, 1980]. The magnitude of the desired frequency response of the CISIM filter was thus set to decay exponentially with increasing distance from the stimulating electrode. The space constant of the exponential decay λ and the position of the simulated electrode $x_{electrode}$ are two free parameters. Consequently the desired frequency response of the CISIM filter is given by:

$$F(f) = e^{-\frac{\left|x(f) - x_{electrode}\right|}{\lambda}} \tag{6.1}$$

The conversion of distance along the cochlea to frequency $x(f)$ was done using Greenwood's formula [Greenwood, 1990]. The desired frequency response was calculated for 400 frequencies spaced equidistantly from 1 mm to 34 mm, assuming

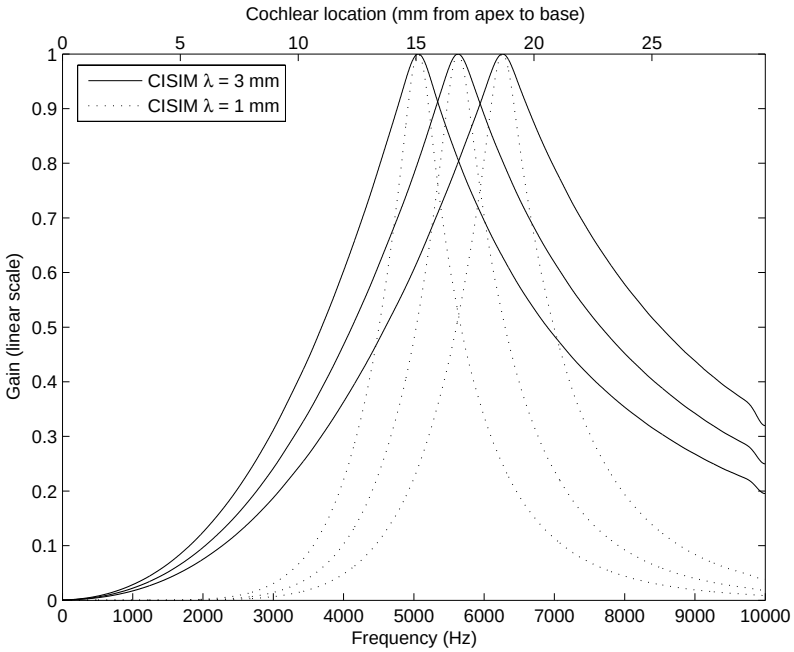


Figure 6.2: Frequency response of six CISIM filters. CISIM filters were generated for three center frequencies (or electrode locations $x_{electrode}$) and for two space constants λ. The cochlear locations of the simulated electrode positions $x_{electrode}$ were 25 mm, 25.75 mm, and 26.5 mm, corresponding to center frequencies 5065 Hz, 5636 Hz, and 6269 Hz. The space constants λ were 3 mm or 1 mm.

a 35 mm cochlea. The obtained desired frequency response was approximated using a 150 tap linear phase FIR filter using the FIR2 command in Matlab.

The CISIM filters are band-pass filters with a triangular-shaped frequency response when plotted on a log-log scale, because the spacing of characteristic frequencies is approximately logarithmic [Greenwood, 1990]. Taking into account that the characteristic frequency doubles every 4.5 mm along the cochlea, the slope of the edges of the CISIM frequency response is approximately $39/\lambda$ dB/octave. As an example, Figure 6.2 shows the frequency response for 6 different CISIM filters.

6.3 Experiment 1A: Simulated place pitch discrimination

In this and the following section two experiments examine the factors affecting pitch perception using noise-band vocoders in normal-hearing subjects for mimicked place pitch and temporal pitch, respectively. These experiments thus investigate the parameters of the resynthesis stage of a vocoder and their influence on pitch perception. The first experiment (section 6.3) assesses the effect of center frequency

and bandwidth of noise bands on place pitch sensitivity by measuring center frequency discrimination, simulating electrode discrimination. The second experiment (section 6.4) studies the effect of these same factors on temporal pitch sensitivity or modulation rate discrimination.

In this experiment, simulated place pitch discrimination was measured using the acoustical analogy of electrode discrimination based on a noise-band vocoder: center frequency discrimination for noise bands. The center frequency discrimination was assessed as a function of the slope of the noise bands and at two different center frequencies.

6.3.1 Methods

6.3.1.1 Subjects

Five NH subjects (NH1 through NH5) aged between 24 and 39 participated in this and all further experiments of this chapter. All subjects were members of the departmental staff and included the author of this thesis. All subjects had experience in psychophysical studies with similar stimuli.

The results of the normal-hearing subjects will be compared to the results of four post-lingually deafened CI subjects that performed equivalent tasks. The results of the CI subjects have been reported in chapter 5. All CI subjects were implanted with the Nucleus CI24 implant and were relatively good performers. Some of the subjects' relevant details can be found in Table 3.1 in chapter 3.

6.3.1.2 Stimuli

Stimuli consisted of white noise filtered with a custom designed CISIM filter (section 6.2.2) which was also used in chapter 4. CISIM filters were designed for each center frequency (electrode position) and each space constant tested.

Center frequency discrimination was measured at two reference center frequencies 565 and 2643 Hz (corresponding to cochlear locations of 10.75 mm and 20.5 mm; cochlear length is expressed from apex to base and assuming a 35 mm cochlea), and for ten different space constants: λ = 0.5, 0.67, 1, 1.33, 2, 2.5, 3.33, 5, 10, and 20 mm. Stimuli were 500 ms long including 10 ms linear on- and off-ramps and were sampled at 20 kHz.

6.3.1.3 Psychophysical procedure

Discrimination was measured using a 2-Interval 2-Alternative Forced Choice (2I-2AFC) constant stimuli paradigm. The subjects were presented with a reference and a comparison stimulus in random order, and they were asked to indicate the highest in pitch. The center frequency of the reference stimulus corresponded to one of the two reference center frequencies tested. The center frequency of the comparison stimulus was higher than that of the reference stimulus, corresponding to a shift towards the base of the cochlea. For space constants λ of 2 mm up to 20 mm three comparison stimuli were used with cochlear locations (corresponding to their center frequencies) shifted 0.75, 1.5 or 2.25 mm towards the base from the cochlear

location of the reference stimulus. This corresponds to shifts of 1, 2, or 3 electrodes in the Nucleus CI24 device. For space constants λ of 0.5 up to 1.33 mm, five comparison stimuli were used and the shifts were 0.1875, 0.375, 0.5625, 0.75, or 1.5 mm towards the base. This corresponds to shifts of 1/4, 1/2, 3/4, 1, or 2 electrodes in the Nucleus CI24 device. Note that a shift of 0.75 mm corresponds to a relative change of center frequency of approximately 12% and that the relative center frequency is roughly proportional to the shift. The intensity of the stimuli was randomly attenuated over a 20dB range to minimize loudness cues.

The trials were presented in blocks. Each block contained trials for a given space constant and for both reference center frequencies. Every trial was repeated ten times per block, and this led to 60 or 100 trials per block for space constants greater or smaller than 2 mm respectively. Every block was presented twice to the subjects; hence every subject was presented with 20 blocks.

In this and all further experiments of this chapter, the acoustical stimuli were routed via a desktop PC and a 24-bit PCI sound card (Lynx ONE, Lynx Studio Technology), to a mixer-amplifier (Eurorack MX1604A) and played to the right earpiece of a TDH39 headphone. The APEX program (see chapter 2) controlled stimulus playback and response collection. The attenuated intensity of the stimuli was adjusted to be at a comfortable level, which in most cases agreed with 70 dB SPL. The subjects were seated in a double-walled soundproof booth.

6.3.1.4 Data analysis

Just noticeable differences (JND) of simulated electrode position were estimated for each combination of reference center frequency and space constant. Using a non-linear least squares fitting routine, a normal cumulative distribution function was fitted to the average proportion of trials where the comparison stimulus was indicated as higher than the reference stimulus as a function of distance between cochlear locations of comparison and reference stimulus. The JND was set to the distance that resulted in a proportion of 75% correct according to the fitted curve. If the listener could not perform the task the discrimination threshold was arbitrarily set to 75 mm for the statistical analysis. This procedure is identical to that used in chapter 3.

6.3.2 Results

Figure 6.3 shows the average JNDs (expressed in mm) of the subjects as a function of the space constant for both reference center frequencies. The logarithms of the JNDs were used for averaging and to obtain a normal distribution for the statistical analysis.

A repeated measures ANOVA was performed on the logarithms of the estimated JND data with the reference center frequency and the space constant as factors. There was a significant effect of space constant (p = 0.002). No other effects were significant.

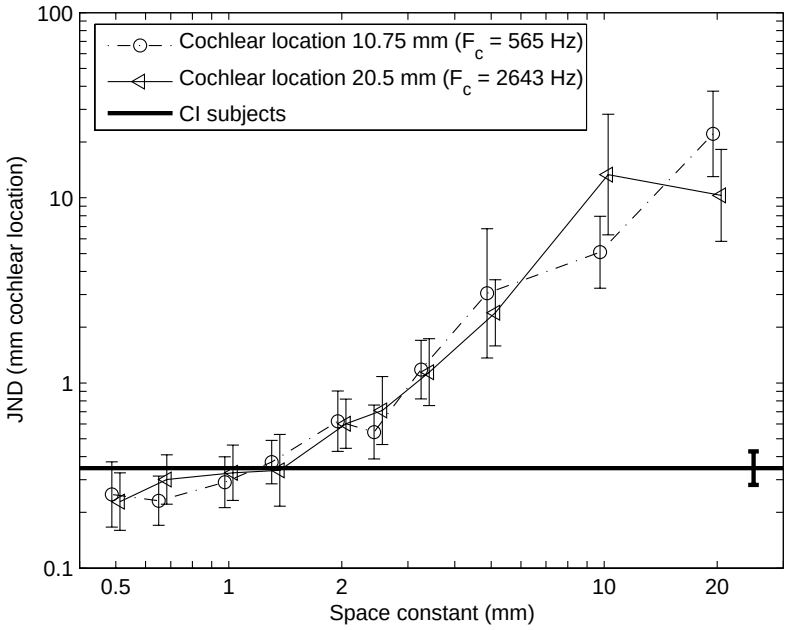


Figure 6.3: Results for simulated electrode discrimination using the CISIM filter. Electrode discrimination was simulated by discrimination of center frequency of noise bands. Just Noticeable differences were estimated as a function of the spectral overlap between the noise bands and expressed as number of electrodes spaced 0.75 mm apart. As the space constant decreases, the noise bands become narrower with less overlap, and the discrimination improves. The error bars indicate ± 1 standard error of the mean of the inter subject variability. There was no significant difference between the two simulated insertion depths. The average result of electrode discrimination of four Nucleus CI24 subjects as measured in chapter 3 is added with the horizontal solid line. The standard error of the mean for the CI subjects is approximately 0.05 mm.

6.3.3 Discussion

The variation of JND as a function of space constant is qualitatively equivalent to the resonance frequency discrimination of filtered noise [Gagne and Zurek, 1988]. Gagne and Zurek found that when the Q-factor of the noise-shaping filter was decreased, the center frequency discrimination was impaired. Similarly in the present experiment, increasing the space constant or broadening the spectral band resulted in an increase of place pitch JNDs. Gagne and Zurek also found that the difference limen expressed relative to the center frequency is roughly constant when the Q-factor or spectral slope was held constant. This is also found in the present study in that there is no difference between the results obtained at both reference center frequencies and noting that a shift in cochlear position corresponds to a constant relative frequency difference.

The JNDs measured by Gagne and Zurek [1988] are however significantly smaller than the JNDs obtained in the present study, especially for large space constants (In comparing the JNDs the d' values derived from the JNDs of both studies were assumed to vary linearly with distance along the basilar membrane). For the larger space constants it is however hard to compare the two filter shapes, since for low values of Q the resonance filters of Gagne and Zurek become low-pass filters with a steeper slope compared to the slopes of the filters used in this experiment. Part of the lower sensitivity in the present study is also most probably due to the relatively large range of intensity roving we applied while Gagne and Zurek did not rove the intensity of their stimuli. Also differences between subjects of the two studies might have contributed to the lower sensitivity reported in this chapter. In the present study a large variability in subjects' performance is observed (see Figure 6.3 and below) which is mainly caused by the performance of two NH subjects performing relatively poor on pitch discrimination (see Figure 6.12).

Using the present custom-made CISIM filter there was no difference in place pitch sensitivity across reference center frequency, or equivalent simulated electrode position. Similarly, in cochlear implants place pitch sensitivity is not significantly different at different electrode locations [Nelson et al., 1995].

The place pitch discrimination of the CI subjects was described in chapter 3. Electrode discrimination was measured for single and multi-channel stimuli. The experimental procedure was very similar to the procedure used for the present simulated electrode discrimination experiment. In the cochlear implant experiment it was however impossible to change the space constant λ of the decay of the excitation pattern. The average JND for electrode discrimination of the four CI subjects is added as a solid horizontal line to Figure 6.3 for comparison. The average performance of the normal-hearing subjects matches the average CI performance for a space constant of approximately 1 mm. Note that the filter designed based on a space constant of 1 mm, has relatively steep slopes of approximately 40 dB/octave.

6.4 Experiment 1B: Simulated temporal pitch discrimination

6.4.1 Methods

Modulation rate discrimination for noise bands was measured as a function of both center frequency and steepness of the spectral slopes. The subjects and procedure for presentation were identical to the experiment in section 0.

Modulation rate discrimination was measured in a 2I-2AFC constant stimuli paradigm. The five NH subjects were presented with two modulated noise bands and were asked to indicate the highest in pitch. One stimulus (the reference) was always modulated at 133 Hz while the comparison stimulus was modulated at a higher frequency (8 frequencies between 135 and 250 Hz, corresponding to frequency differences from 0.26 to 10.9 semitones). On each trial the reference and comparison stimulus had the same spectral shape as both were filtered with the same CISIM filter (same insertion depth and space constant). No feedback was given during the

experiment and the intensity of the stimuli was roved over a 20 dB range to minimize loudness cues.

The stimuli were all derived from one token of 500 ms 'frozen' (i.e. the noise token was fixed) white noise sampled at 20 kHz, which was filtered with the CISIM filter. The noise bands were created for three cochlear positions $x_{electrode}$ of 10.75 mm, 15.25 mm, and 20.5 mm, which correspond to center frequencies of the noise bands of 565, 1194, and 2643 Hz, respectively. Each electrode position was also tested for four different space constants λ 0.5 mm, 1 mm, 2 mm, and 3.33 mm.

The filtered noise bands were sinusoidally modulated with 100% modulation depth at modulation frequencies of 133, 135, 137, 140, 150, 165, 180, 210, and 250 Hz. The stimuli were linearly gated with 10 ms ramps and set to equal RMS power.

The trials were presented in blocks and each trial was repeated four times within a block. The space constant was kept constant within one block but stimuli at the three center frequencies were presented in each block. This led to 24 different trials and hence a total of 96 trials per block. Every block was presented three times to the subjects.

A cumulative normal distribution was fitted to the proportion of correctly ranked trials as a function of the modulation frequency difference expressed as semitones. The JND of modulation rate discrimination was estimated from this curve by taking the frequency difference that corresponds to a proportion of 75% correct.

6.4.2 Results

Figure 6.4 shows the average JND of modulation rate discrimination for the five subjects as a function of the space constant and at the three center frequencies. The logarithms of the JNDs were analyzed with a repeated measures ANOVA with degrees of freedom adjusted to lower bound. Both the center frequency (p = 0.041) and the space constant (p = 0.023) had a significant effect on the modulation rate discrimination performance. There was no significant interaction effect. Both the increase of center frequency (shallower simulated electrode positions) and the increase of the bandwidth of the noise band (larger space constants) improved the modulation rate discrimination. A post-hoc pair-wise comparison using the Least Significant Difference showed that the deepest electrode position differed significantly from the two shallower insertion depths (p = 0.005 and p = 0.049, respectively). However, there was no significant difference between the two shallower electrode positions corresponding to center frequencies 1194 Hz and 2643 Hz (p = 0.590).

Subjects differed strongly in their ability to discriminate modulation rates. The average JND of the poorest performing subject NH3 is approximately 8 times larger than the average JND of the best performing subject NH2. Previous studies have also reported different sensitivities for modulation rate discrimination across subjects [Patterson et al., 1978]. In this study, for one subject the modulation depth had to be up to 4 times deeper than for the other two subjects to obtain the same level of discrimination between two modulation rates. More recently, it has also

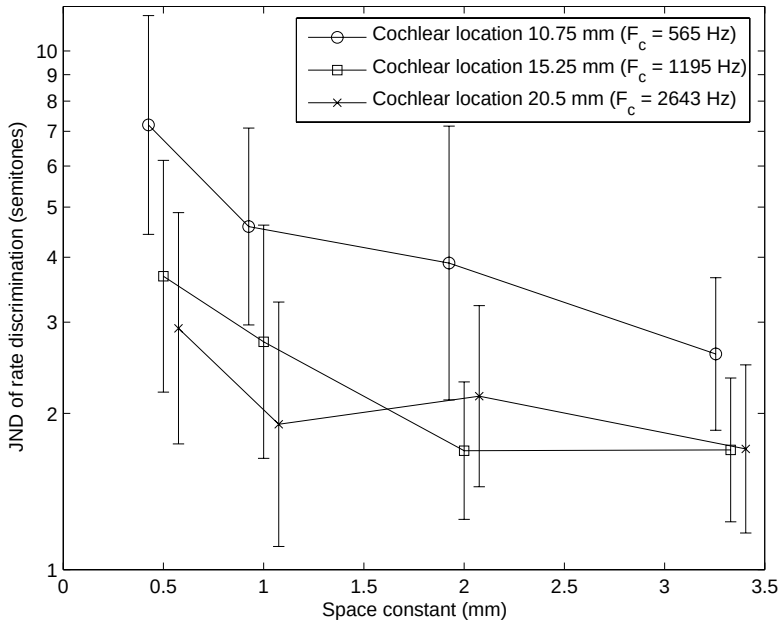


Figure 6.4: Just noticeable differences for modulation rate of noise bands for 5 NH subjects. The noise bands were varied in center frequency and bandwidth by varying the cochlear location and space constant of the CISIM filter respectively. Generally, noise bands with the frequency response extending up to higher frequencies (either through a more basal cochlear location or larger space constants) allow for better modulation rate discrimination. The error bars indicate ± 1 standard error of the mean of the inter subject variability.

been shown that the saliency of envelope pitch varies greatly across a group of 23 normal-hearing subjects [Renken et al., 2004].

6.4.3 Discussion

When the simulated electrode is inserted more deeply or when the space constant becomes larger, the noise band extends to higher frequencies. Consequently, as the upper cut-off frequency of the noise band increases, modulation rate discrimination improves. The results are consistent with the data of Hanna [1992]. He found that for modulation rates above 100 Hz, modulation rate discrimination is limited by peripheral filtering. The basilar membrane starts oscillating at its characteristic frequency when excited with noise and at lower characteristic frequencies these oscillations can obscure the modulation of the noise band. Also, the limited bandwidth of the peripheral filters at lower frequencies reduces the modulation depth of the imposed modulation reducing the discriminability of the modulation rates.

In CI subjects rate discrimination performance also varies over a wide range across subjects. Difference limens for pulse rates between 100 and 150 pulses per second range from 3.1% [McDermott and McKay, 1997] in one musically trained subject, up to 50% in other subjects [Townshend et al., 1987]. This corresponds to frequency differences of 0.5 semitones up to 7 semitones. The range of modulation rate JNDs from our results of simulated modulation rate discrimination overlaps with the range of JNDs found in CI subjects, except for narrow noise bands with relatively low center frequencies.

6.5 Experiment 2: F0-discrimination using noise-band vocoders

6.5.1 Methods

In this section F0-discrimination of stylized synthetic vowels was measured in normal-hearing subjects for two different noise-band vocoders. The subjects and equipment for stimulus presentation were the same as in section 6.3 . The results obtained with each vocoder are compared to the results for CI subjects from chapter 5.

6.5.1.1 Stimuli

The stimuli consisted of stylized vowels processed with a noise-band vocoder. The unprocessed stylized vowels were a subset of the original stimuli used in chapter 5. Stylized vowels were generated by passing a pulse train through a low-pass filter and a single resonator. The low-pass filter was a second order IIR filter with a cut-off at 50 Hz and the output of the low-pass filter resembled the glottal volume velocity. The resonator was a second order IIR resonating filter that created a formant in the spectrum. The details of both filters can be found in section 5.2.3. The formant frequency was set at 300 Hz, 350 Hz, 400 Hz, 450 Hz, or 500 Hz while the bandwidth of the resonator was kept fixed at 100 Hz.

These Single Formant Stimuli (SFS) were created with fundamental frequencies 133 Hz, 140 Hz, 150 Hz, 165 Hz, 180 Hz, 210 Hz, 250 Hz, 325 Hz, and 450 Hz. The sampling frequency was 16 kHz. All stimuli were 500 ms long and had equal RMS power.

6.5.1.2 Sound processing – Analysis stage

An overview of the processing blocks of the vocoders, consisting of an analysis stage and a resynthesis stage, is shown in Figure 6.1. The analysis part, shown on the left hand side of Figure 6.1, is identical for both noise-band vocoders tested in the present experiment and is identical to the analysis stage of the processing in chapter 5. Four different filter banks were used to process the stimuli in the analysis stage. The filter banks differ in spectral resolution for the low frequencies and in the shape of the individual filters. All filter banks have 22 bands and their frequency response is shown in Figure 5.1. One filter bank (ACE) is identical to the filter bank currently implemented in the commercial speech processor of Nucleus devices

[Cochlear Ltd, 2002]. It consists of a 128-point FFT with a Hamming window. The 64 frequency bins are summed together to form 22 channels with an approximately logarithmic frequency resolution. The three other filter banks resemble more closely the frequency analysis of the normal human ear and have more resolution in the low frequencies compared to the ACE filter bank. The GTF filter bank is based upon a model for the filtering of the normal ear [Patterson et al., 1995]. The BUT and GTFW are derived from this filter bank by changing the filters' shape and band width, respectively. The GTF, BUT and GTFW filter bank all consist of 22 4th order IIR band-pass filters with center frequencies spaced approximately equidistantly along the length of the basilar membrane. The filters in the GTF and GTFW filter banks were 4th order gammatone filters, while Butterworth filters were used for the BUT filter bank. The bandwidth of the GTF filter bank was set to the ERB of the respective auditory filter [Glasberg and Moore, 1990] whereas the bandwidth of the GTFW filter bank was increased so that all overtones were resolved. For the BUT filter bank the bandwidth was set to have cross-over frequencies at 3 dB.

In order to separate the effects of temporal and place pitch cues, the envelopes of the filter outputs were obtained under two different conditions. In one condition, including both temporal and place pitch cues, the envelope contained modulations up to 200 Hz. In the other condition, including place pitch cues but without temporal pitch cues, the envelope was low-pass filtered at 10 Hz so that all fluctuations causing temporal pitch cues were removed. More details of the implementation of the four filter banks and the envelope extraction are given in chapter 5.

Finally, maxima selection was performed for all processing schemes in two steps. Firstly, the signals were resampled from the audio sampling rate to 900 Hz (the stimulation rate per channel in the CI). Then, for each time slot (sampled at 900 Hz) only the 8 most intense channels of the 22 channels were retained, while the other channels were set to zero. In contrast to CI-processing no compression of the envelopes was included in the noise-band vocoder schemes.

6.5.1.3 Sound processing – Resynthesis stage

The specification of the noise bands, serving as carriers in the resynthesis stage of the noise-band vocoders, differs between the two vocoders: standard vocoder and CISIM vocoder.

a. Standard vocoder

In the standard vocoder the output of the analysis stage is upsampled to 16 kHz using sample-and-hold resampling and then modulated with noise bands. These noise bands are generated by filtering white Gaussian noise through 4^{th} order Butterworth band-pass filters. The cut-off frequencies of the resynthesis filters are set equal to the cross-over frequencies of the filters of the analysis filter bank. Consequently, the resynthesis filters differ depending on the filter bank used in the analysis.

For the BUT and ACE filter bank the cross-over frequencies of the analysis filter bank were determined analytically. For the GTF filter bank the cut-off frequencies of the resynthesis filter bank were set to 23 frequencies spaced evenly along the

basilar membrane between 100 and 8000 Hz, where the conversion between place along the basilar membrane and frequency was calculated using Greenwood's formula [Greenwood, 1990]. The standard vocoder was not used with the GTFW filter bank.

b. CISIM vocoder

The main difference of the CISIM vocoder with the standard vocoder was the frequency response of the resynthesis filters. The CISIM vocoder used 22 CISIM filters (defined in section 6.2.2), to simulate an electrode array consisting of 22 electrodes. The cochlear locations of the center frequencies of the 22 CISIM filters were spaced 0.75 mm apart similar to the electrode spacing of the Cochlear Nucleus CI24(M) electrode array. The space constant of the 22 CISIM filters was set to 1 mm, because for this value the performance of the simulated electrode discrimination in normal-hearing subjects in section 6.3 was closest to the electrode discrimination of the CI-subjects (see above). The insertion depth of the deepest electrode of the simulated electrode array was set relatively shallow to 20 mm from the round window. This was done to minimize the negative effect of the peripheral auditory filters on temporal pitch sensitivity (see section 6.4). In contrast to the standard vocoder, the resynthesis filters were now kept fixed independent of the analysis filter bank. This resembles more closely CI operation because when the sound processing scheme is altered, the channels perceived by the subject do not immediately change accordingly because the electrodes remain at the same location and the current spread does not alter. It is noted however that after some time CI subjects do tend to adapt to new schemes [Fu et al., 2002]. Such adaptation is not active in the experiments in chapter 5 because the experiments were acute.

6.5.1.4 Psychophysical procedure

F0-discrimination was measured in a 2I-2AFC procedure using the method of constant stimuli. The subjects were presented with two stimuli separated by a 500 ms silent gap and were asked to indicate which of the two intervals contained the stimulus with higher pitch, while ignoring the loudness cues. No feedback was presented to enforce subjects to use their intuitive sense of pitch and to prevent the subjects from using other cues than pitch. Both stimuli in a trial had the same formant frequency but one stimulus' F0-frequency was set at the reference frequency (133 Hz) and the F0 of the other stimulus was set at a higher frequency.

The trials were presented in blocks. The conditions resulting from the combinations of type of envelope extraction (with or without temporal pitch cues) and filter bank (ACE, GTF, BUT, or GTFW), were presented into separate blocks. Every block was presented twice to every subject. Each block contained trials with every comparison frequency for all 5 formant frequencies, and within each block each trial was repeated twice. Different formant frequencies were included per block in order to prevent subjects to identify and learn the reference stimulus based upon other cues.

For the standard vocoder the F0 of the comparison stimuli only ranged up to 250 Hz and the GTFW filter bank was not included, while for the CISIM vocoder the fundamental frequency of the comparison stimuli ranged up to 450 Hz, as in chapter 5. In summary, 12 blocks of 60 trials containing stimuli processed with the standard

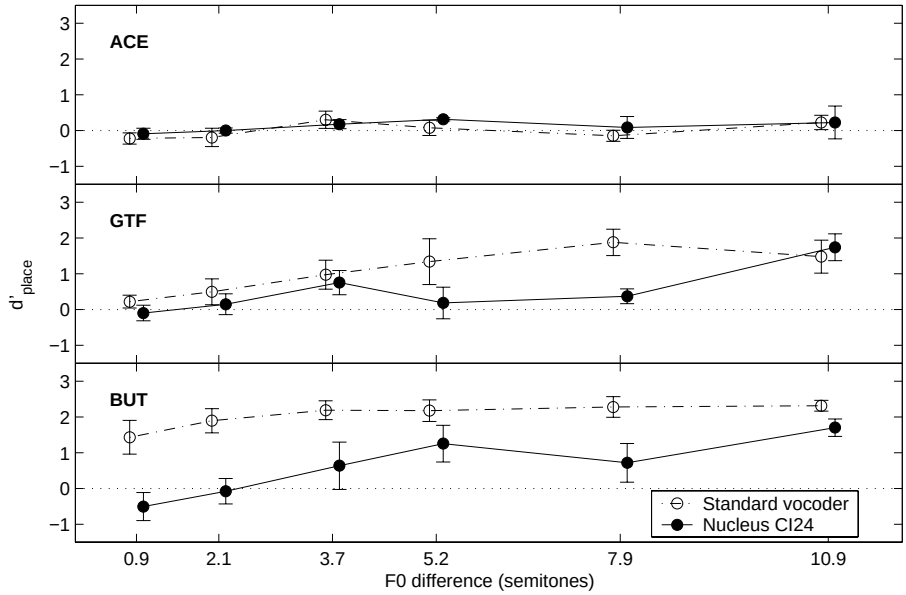


Figure 6.5: The average the effect of place pitch cues for F0 discrimination present in the stimuli processed with the standard vocoder for the ACE, GTF and BUT filter banks. The results are averaged over the five NH subjects. The average result for four CI subjects from chapter 5 is included for comparison. The error bars indicate ± 1 standard error of the mean of the inter subject variability.

vocoder and 16 blocks of 80 trials containing stimuli processed with the CISIM vocoder were presented to every subject.

To minimize loudness cues, the intensity of the stimuli was randomly attenuated over a 20 dB range.

6.5.1.5 Data analysis

A measure of the effect of temporal and place pitch cues was derived from the proportions of correctly ranked trials. Because no significant effect of the formant frequencies was found in chapter 5 and in the present data, the proportions were averaged over the different formant frequencies. The resulting average proportions were transformed into *d'* values, taking into account an adjustment for perfect performance. For all proportions equal to one, half a trial error was introduced as suggested by MacMillan and Creelman [1991].

The d'_{place} values were estimated from the condition where the temporal fluctuations in the envelopes above 10 Hz were filtered out and represent an estimate of the effect of place pitch cues. The $d'_{combined}$ values were estimated from the condition where temporal fluctuations in the envelopes were present up to 200 Hz, and represent an estimate of the combined effect of temporal and place pitch cues. The effect of the pure temporal pitch cues (d'_{temp}) was calculated by taking the difference

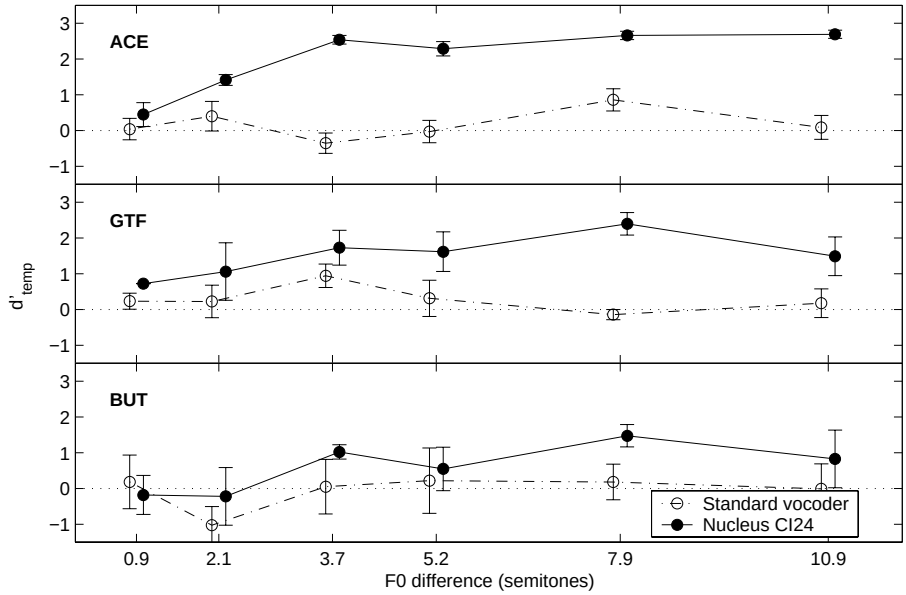


Figure 6.6: The effect of the temporal pitch cues for F0 discrimination for stimuli processed with the standard vocoder for different filter banks and as a function of relative F0 difference. The results are averaged over the five NH subjects. The error bars indicate ± 1 standard error of the mean of the inter subject variability. The average result for four CI subjects from chapter 5 is included for comparison.

between the effect of the combined pitch cues and the effect of the place pitch cues using Eq. (5.20).

In order to compare the results obtained with the vocoder and the results obtained for CI subjects, the *d'* values for place pitch cues, temporal pitch cues and combined temporal and place pitch cues were calculated for the results of the normal-hearing subjects using the vocoders and for the results of 4 CI subjects from chapter 5. For both vocoders, a repeated measures ANOVA was performed on the *d'* values with two within-subject factors (analysis filter bank and F0-difference) and one inter-subject factor (separating NH-subjects from CI-subjects). This analysis was performed for d'_{place}, d'_{temp}, and d'_{combined} separately.

6.5.2 Results

6.5.2.1 Standard vocoder

Figure 6.5, Figure 6.6, and Figure 6.7 display the effects of place pitch cues, temporal pitch cues and combined pitch cues, respectively, obtained with the standard vocoder. Each panel shows results of stimuli processed with a different analysis filter bank. The average results for the 4 CI subjects from chapter 5 are included in the figures for comparison.

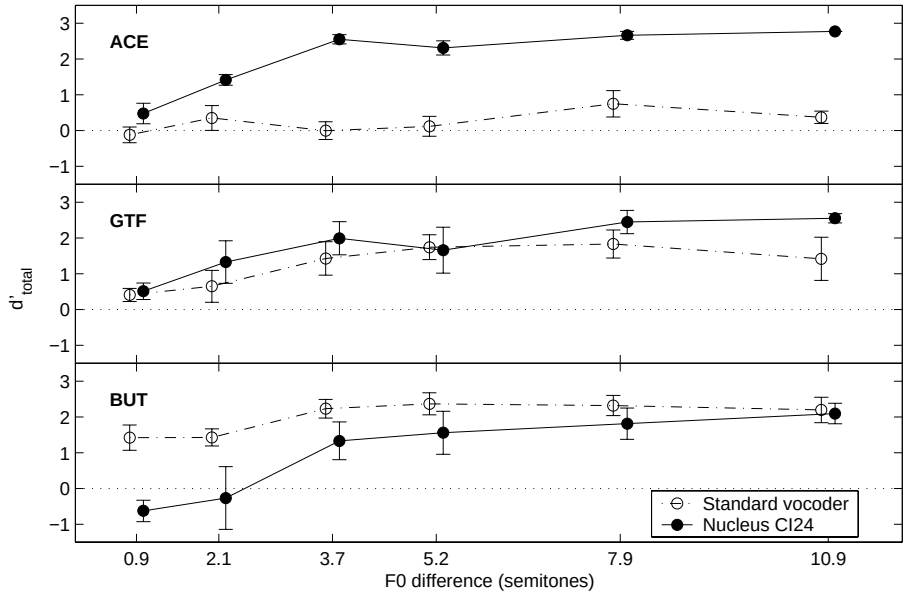


Figure 6.7: The effect of combined temporal and place pitch cues for F0 discrimination for stimuli processed with the standard vocoder for different filter banks an as a function of F0 difference between reference and comparison. The results are averaged over the five NH subjects. The error bars indicate ± 1 standard error of the mean of the inter subject variability. The average result for four CI subjects from chapter 5 is included for comparison.

In the following paragraphs the effects of the place-pitch, temporal-pitch and combined-pitch cues are analyzed separately.

- The repeated measures ANOVA performed on the d'_{place} values (shown in Figure 6.5) comparing the vocoder results with the CI-results, indicate that there was a significant difference between the results obtained with the vocoder and the results obtained with the CI subjects (p = 0.048). Moreover, this difference depended significantly on the analysis filter bank (p = 0.042). The place pitch cues had a significantly larger effect in the NH-subjects than in the CI-subjects. The largest difference between the standard vocoder results and CI results occurred for the BUT filter bank, where the standard vocoder allowed the subjects to discriminate frequency differences smaller than a semitone (approximately 5%) solely based upon place pitch cues. For the ACE filter bank the place pitch cues had almost no effect both with the standard vocoder and with the CI.
- The analysis of the d'_{temp} values (shown in Figure 6.6) indicated that the temporal pitch cues had significantly different effects for the NH-subjects with the standard vocoder than for the CI-subjects (p = 0.002). The perceived pitch effect of adding temporal modulations in the envelope up to 200 Hz was smaller for the NH-subjects with the standard vocoder than for the CI-subjects.

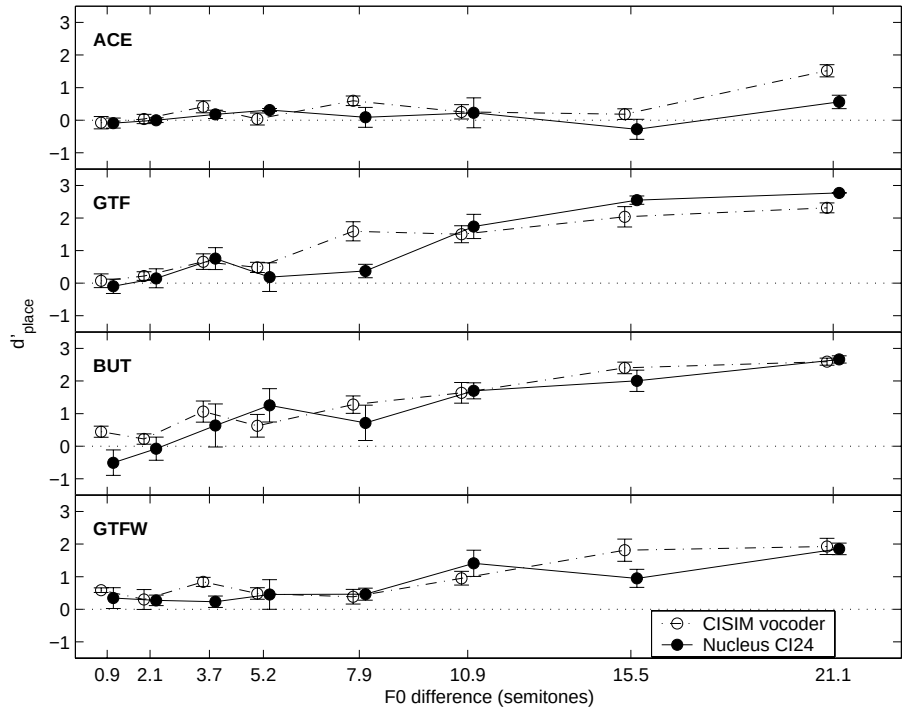


Figure 6.8: The effect of the place pitch cues for F0 discrimination for stimuli processed with the CISIM vocoder for different filter banks and as a function of relative F0 difference. The results are averaged over the five NH subjects. The error bars indicate ± 1 standard error of the mean of the inter subject variability. The average result for four CI subjects from chapter 5 is included for comparison.

- The combined effects of both temporal and place pitch cues ($d'_{combined}$ shown in Figure 6.7) were also compared between the standard vocoder and the CI. Although, there is no significant difference for F0-discrimination with combined place and temporal pitch cues between the CI subjects and normal-hearing subjects with the standard vocoder over all the filter banks ($p = 0.295$), there is a significant difference depending on the processing filter bank used ($p = 0.001$). For example, the ACE filter bank led to the best performance in the CI subjects but to the worst performance in the normal-hearing subjects using the standard vocoder. Similarly, while the BUT filter bank led to the best performance in normal-hearing subjects, it had the worst performance of all filter banks in the CI subjects.

6.5.2.2 CISIM vocoder

Figure 6.8, Figure 6.9, and Figure 6.10 show the effect the place pitch cues (d'_{place}), the effect of the temporal pitch cues (d'_{temp}), and the effect of the combined place and temporal pitch cues ($d'_{combined}$) respectively averaged over the 5 NH-subjects for the CISIM vocoder. Each panel shows the results for a different analysis filter bank.

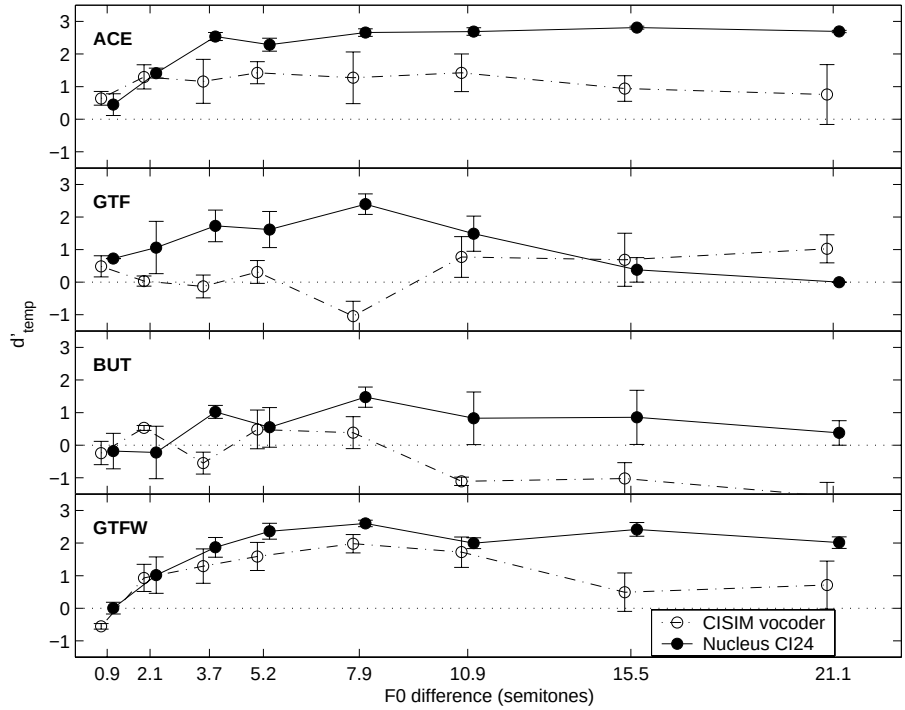


Figure 6.9: The effect of the temporal pitch cues for F0 discrimination for stimuli processed with the CISIM vocoder for different filter banks and as a function of relative F0 difference. The results are averaged over the five NH subjects. The error bars indicate ± 1 standard error of the mean of the inter subject variability. The average result for four CI subjects from chapter 5 is included for comparison.

The average results for the 4 CI-subjects from chapter 5 are included in each figure for comparison.

In the following paragraphs the effects of the place-pitch, temporal-pitch and combined-pitch cues are analyzed separately.

- The statistical analysis of the perceptual effect of the place pitch cues (d'_{place}), showed no significant difference between the CISIM vocoder and the CI results (p = 0.412). There were significant effects of the filter bank (p = 0.002), the relative F0 difference (p < 0.001), and a significant interaction effect between the filter bank and the relative F0 difference (p = 0.023).
- The analysis of the amount of effect of temporal pitch cues (d'_{temp}) showed that both the filter bank and the relative F0 difference had a significant effect on performance (p = 0.002, and p = 0.032, respectively). There was a significant difference between the results obtained with the CISIM vocoder and with the CI subjects (p = 0.028). However the procedure for obtaining these results (with the CISIM vocoder or with CI subjects) did not significantly interact with any term.

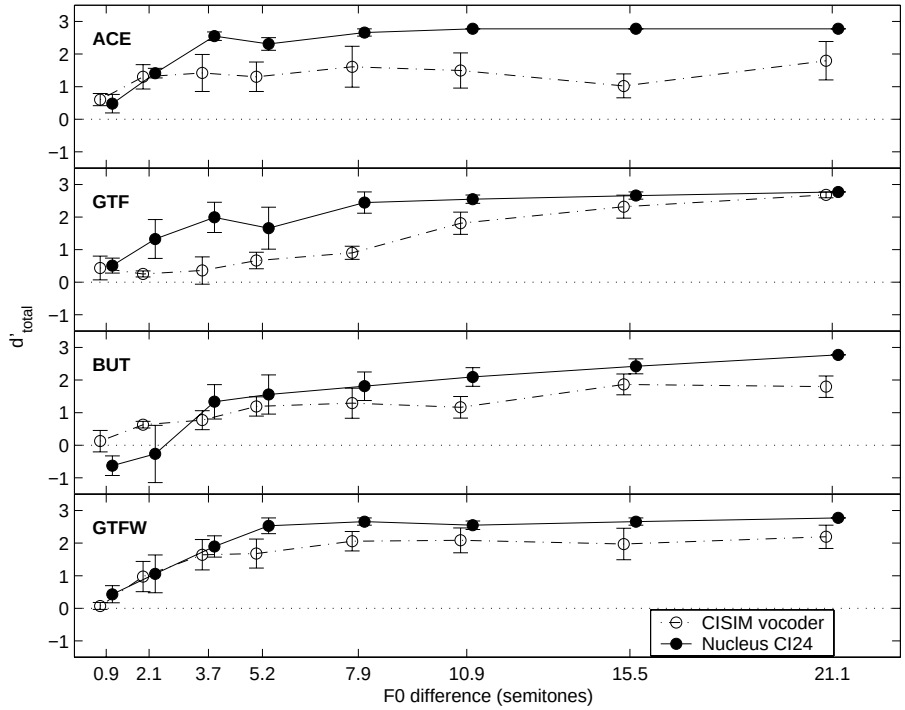


Figure 6.10: The effect of the combined place and temporal pitch cues for F0 discrimination for stimuli processed with the CISIM vocoder for different filter banks and as a function of relative F0 difference. The results are averaged over the five NH subjects. The error bars indicate ± 1 standard error of the mean of the inter subject variability. The average result for four CI subjects from chapter 5 is included for comparison.

- The analysis of the effects of the combined place and temporal pitch cues (d'_{combined}) showed only a significant effect of the relative F0 difference ($p < 0.001$). There was no significant overall difference between the CISIM data and the results obtained from the CI subjects ($p = 0.140$).

The CISIM vocoder performed well in modeling the effect of the place pitch cues present in the CI results. There is approximately no difference between the results obtained from NH-subjects with the CISIM vocoder and from CI-subjects for the condition where only place pitch cues were present. Moreover, the comparison in performance over the four different analysis filter banks leads to similar results when using the CISIM vocoder results or the CI results: The effect of the place pitch cues was largest for the BUT and GTF filter banks, while for the ACE filter bank there was almost no effect of place pitch cues.

However, the effects of the temporal pitch cues are smaller with the CISIM vocoder than those present in the CI results. For both the ACE and GTFW filter bank F0-discrimination performance improved by adding temporal modulations in the envelopes up to 200 Hz. Temporal pitch cues thus had an effect for these filter banks, although the effect is smaller than the effect found in the CI subjects. It

should also be noted that the inter subject variability is larger for the CISIM vocoder than for the CI results. For the GTF and BUT filter bank no clear benefit in F0-discrimination is obtained by providing temporal modulations in the envelopes up to 200 Hz for the CISIM vocoder. Even more, for the BUT filter bank for large F0-differences the F0-discrimination performance dropped after adding the temporal fluctuations.

In summary, the effect of the place pitch cues is approximately equal for the CISIM vocoder and the CI but the effect of the temporal pitch cues is smaller for the CISIM vocoder. Consequently, the effect of the combined place and temporal pitch cues is on average somewhat smaller for the CISIM vocoder for all analysis filter banks, although this effect is not significant.

6.5.3 Discussion

In most noise-band vocoders used for CI research white noise is modulated before the resynthesis filters are applied [Shannon et al., 1995]. In the present study the modulation of the noise is performed after the filtering for two reasons. Firstly, the band pass filtering may reduce the depth of the modulation of the envelope for higher modulation frequencies. And secondly, the different group delays of the resynthesis filters alter the phase relations between the modulations on the different channels.

Although both the standard and the CISIM vocoder can be used to measure F0-discrimination, they are not completely suitable as acoustic models for CI. For the standard vocoder neither the effects of the place pitch cues, nor the effects of the temporal pitch were correctly modeled. For the CISIM vocoder the effect of the temporal pitch cues was somewhat underestimated, but the effects of the place pitch cues were accurately modeled for all analysis filter banks. In the next sections we discuss the possible factors affecting the results for place pitch cues and temporal pitch cues for both vocoders.

6.5.3.1 Place pitch cues

With the standard vocoder the normal-hearing subjects obtained better F0-discrimination than the CI subjects when the discrimination was solely based on place pitch cues. This higher pitch sensitivity using the standard vocoder was most likely caused by the relatively high frequency resolution of the vocoder processed signals. The resynthesis filters of the standard vocoder were relatively steep and did not overlap. Consequently, the spectrum of the signals was relatively little smeared by the standard vocoder. In contrast, excitation patterns of adjacent channels in CI had shallow slopes and overlapped greatly because of the current spread along the cochlea [Shannon, 1983b]. This overlap smeared the resulting excitation pattern and consequently weakened the place pitch differences between the compared signals.

The high performance of the NH subjects with the BUT filter bank using purely place pitch cues was likely due to the fact that for the BUT filter bank the analysis and resynthesis filters were identical. In this way the spectrum was least distorted for the BUT filter bank and the spectrum was minimally warped along the cochlea.

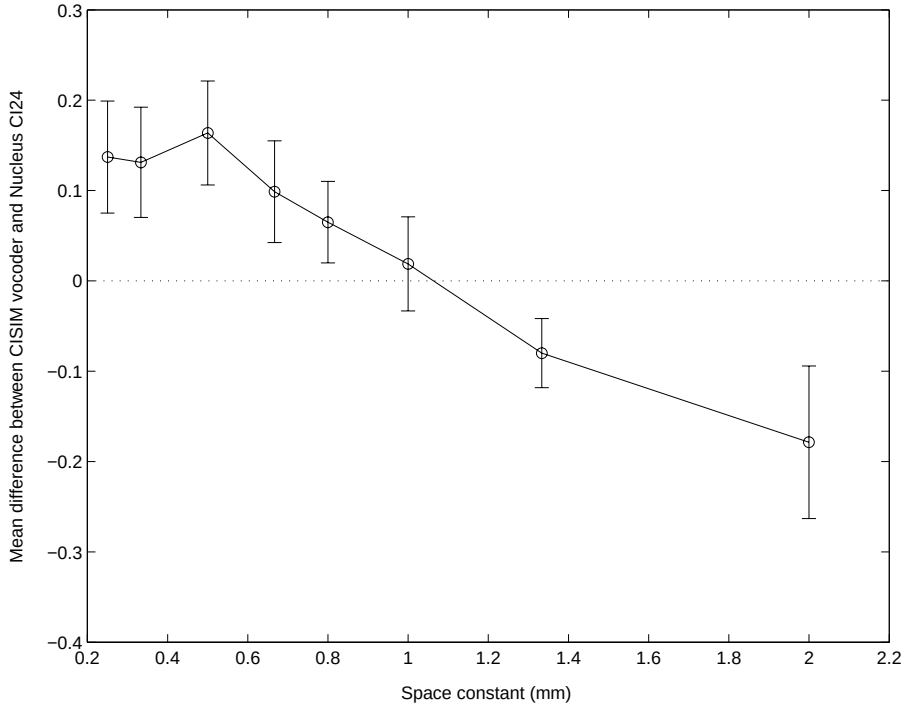


Figure 6.11: The mean difference between the average psychometric curve of the CI subjects and the average psychometric curve of the NH subjects with the CISIM vocoder as a function of the space constant of the CISIM filters. Positive values indicate that the psychometric curves for the CI subjects were on average higher than for the NH subjects, and consequently had better F0-discrimination and vice versa. All stimuli were processed with the GTF filter bank and with the envelopes low-pass filtered at 10 Hz. The best correspondence between the vocoder and CI results is a space constant of 1 mm. The error bars indicate ± one standard error of the mean of the five normal-hearing subjects.

Presenting the correct frequency to the correct place of stimulation is crucial for NH subjects to obtain good frequency discrimination [Oxenham et al., 2004].

For the CISIM vocoder the slopes of the resynthesis filters were set to obtain equal channel discrimination using the CISIM vocoder as found in CI subjects (see section 6.3). With this adjusted channel discrimination, the results of the NH subjects with the CISIM vocoder successfully model the results of CI subjects in an F0-discrimination task based solely on place pitch cues.

To highlight the importance of channel overlap for correctly modeling the CI results, a control experiment was set up. F0-discrimination was measured for the CISIM vocoder with the default insertion depth and with various space constants using the above described stimuli, subjects and procedure but only for the GTF filter bank with the envelopes low-pass filtered at 10 Hz. In this way, the effect of changing the slope of the resynthesis filters (space constants or equivalently changing the channel

discrimination) on place pitch cues for F0-discrimination was assessed. The tested space constants were 2 mm, 1.33 mm, 1 mm, 0.8 mm, 0.67 mm, 0.5 mm, 0.33 mm, and 0.25 mm.

The psychometric curves were averaged across the 5 NH subjects and the difference was calculated between the average psychometric curve of the NH subjects and the average psychometric curve obtained for CI subjects in the same condition from chapter 5. For each space constant the mean difference is calculated and this mean error is shown in Figure 6.11 as a function of space constant. Positive values indicate that the place pitch cues were more salient for the NH subjects using the CISIM vocoder than for the CI subjects. Negative values indicate that the CI subjects were able to discriminate better F0-differences than the NH subjects using solely place pitch cues.

For small space constants, or easily discriminable adjacent channels, the effect of the place pitch cues was bigger for the NH subjects using the vocoder than for the CI subjects. However, by increasing the channel overlap the effect of the place pitch cues decreased and ultimately became less salient in the vocoder than in the CI subjects. The closest correspondence between the results using the CISIM vocoder and the CI results is obtained for a space constant of 1 mm. This is the value for the space constant that was also found to optimally mimic electrode discrimination (see section 0) and that was used in the default CISIM vocoder.

6.5.3.2 Temporal pitch cues

Although amplitude modulated noise is known to elicit pitch [Burns and Viemeister, 1976] and the sensitivity to temporal pitch in CI is similar to the sensitivity of rate discrimination of amplitude modulated noise[Blamey et al., 1984b], the effect of the temporal pitch cues was smaller for the normal-hearing subjects using the vocoders than for CI subjects. This limited temporal pitch sensitivity using the vocoder was most probably caused by a combination of factors. The most important factors are the poor pitch sensitivity of some of the normal-hearing subjects in the present study, the absence of envelope compression/expansion in the vocoder and the ringing of the peripheral auditory filters of the normal ear.

1. Difference limens for modulation rate discrimination around 133 Hz typically are about 6 % [Formby, 1985; Hanna, 1992]. These difference limens are little affected by the frequency region of the noise bands, except when the spectrum of the noise band is band-pass filtered below 1600 Hz the difference limens rise to approximately 12% [Hanna, 1992]. For two of the subjects of the present study (NH1 and NH3) the average difference limens for modulation rate discrimination measured in section 6.4 for noise bands with center frequency 2643 Hz were 21% and 62%, respectively. These difference limens are a lot poorer than previously reported and poorer than the other subjects (average difference limens of the other subjects is 7%). There is no clear reason why these two subjects performed so poorly at the modulation rate discrimination task.
2. The second factor that most probably contributed to the limited effect of the temporal pitch cues using the noise-band vocoders is the absence of any

compression/expansion of the envelopes in the vocoders. In the CI-system the envelopes are compressed to accommodate for the reduced dynamic range and the steep loudness growth of CI subjects. However, even with compression, the resulting modulation depth in the neural excitation pattern may be larger for CI subjects compared to the modulation depth in the excitation pattern of NH subjects listening to the vocoders. This is because excitation is an expansive function of the input for electrical stimulation, while it is a compressive function for acoustic stimulation [Zeng and Shannon, 1994]. Consequently, the relative reduced modulation depth may have impaired the effect of the temporal pitch for the NH subjects because lower modulation depths lead to poorer modulation rate discrimination [Patterson et al., 1978; Grant et al., 1998].

To test this latter hypothesis the minimal modulation depth required to discriminate a 20% change in modulation frequency (approximately 3.2 semitones) was measured in the same five NH subjects using the CISIM vocoder and in the four CI subjects of chapter 5. For the CI subjects, the stimuli consisted of amplitude modulated pulse trains presented interleaved on the three most apical channels with 900 pulses per second per channel. A DC-shifted sinusoid was compressed using the standard compression function to accommodate for loudness growth in Nucleus CI24 subjects (Eq. 5.1) and used to modulate the amplitude of the pulses of the three channels. The modulation was in phase across the three channels and the modulation depth was varied adaptively. The subjects were presented two signals in random order on each trial: one was modulated at 133 Hz and the other at approximately 160 Hz. Subjects were asked to indicate the highest in pitch. After two consecutive correct answers the modulation depth was decreased by 1 dB and after each incorrect answer the modulation depth was increased again by 1 dB leading to an asymptotic average of 71% correct responses [Levitt, 1971]. The procedure was continued until 8 reversals were obtained and the mean of the last four reversals was taken as the result for that particular run. The three best runs out of five runs were retained for each subject. Intensity was roved identically as in chapters 3 and 5 by randomly varying the volume from 85% up to 110% of the dynamic range, to minimize loudness cues. For the normal-hearing subjects, stimuli consisted of the sum of three modulated noise bands filtered with CISIM filters with space constant of 1 mm and insertion depths equivalent to the three most apical electrodes of the simulated electrode array of the CISIM vocoder (20 mm, 19.25, and 18.5 mm respectively, corresponding to center frequencies of 1148 Hz, 1291 Hz, and 1451 Hz). The modulation was sinusoidal and in phase over the three channels. The procedure was identical to the procedure for the CI subjects. Intensity was roved over a 20 dB range. No compression or expansion of the envelope was included for the acoustical stimuli.

Figure 6.12 shows the results for both the NH subjects and the CI subjects. Two of the normal-hearing subjects (NH1 and NH3) were unable to discriminate the rate difference even at 100% modulation depth. CI-subjects CI1, CI2 and CI3 require the smallest modulation depth to discriminate the 20% modulation rate difference. CI-subject CI4 is comparable in performance to the better NH subjects. In general, CI-subjects thus require less modulation depth compared to

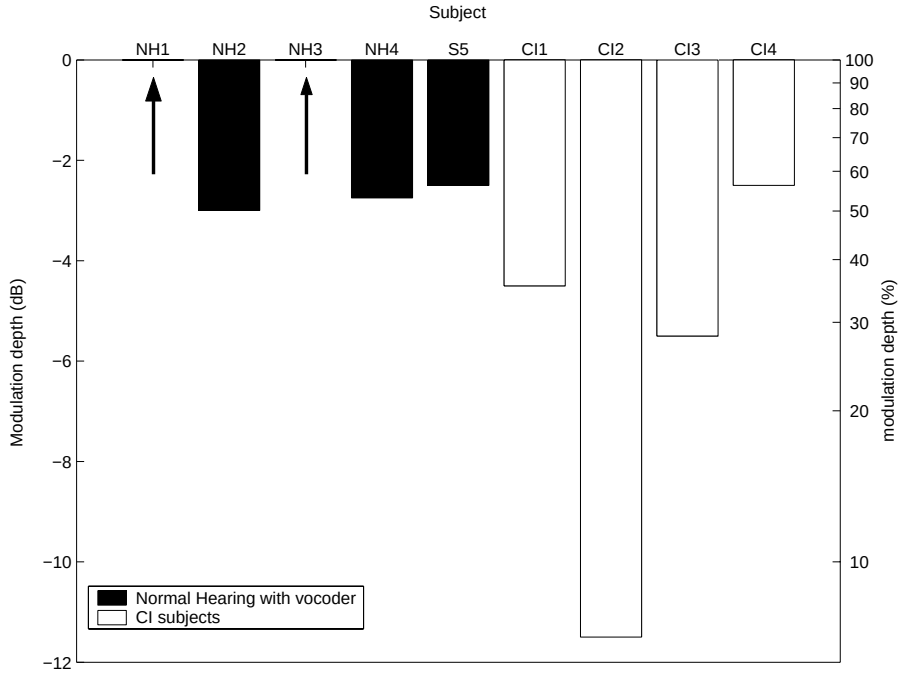


Figure 6.12: Minimal required modulation depth to discriminate a 20% modulation rate difference on three simultaneously stimulated channels. The filled bars show the results of the normal-hearing subjects (NH1-NH5) where the modulation depth was measured using modulated noise bands (CISIM filters with space constant 1 mm). The open bars show the results of the CI subjects (CI1-CI4) where pulse trains were modulated. The arrows for subjects NH1 and NH3 indicate that they were unable to correctly rank the 20% modulation rate difference above chance level even for 100% modulation depth. The results show that the required modulation depth for rate discrimination is generally lower for the CI subjects compared to the normal-hearing subjects.

NH subjects with the CISIM vocoder to discriminate the 20% difference in modulation rate.

The larger modulation depths required for rate discrimination in NH subjects using the vocoders compared to the CI subjects has different effects for the four analysis filter banks. For the ACE and GTFW filter banks the modulation depth present in the reference stimuli ranged from 59% up to 65% for the ACE filter bank and from 35% up to 60% for the GTFW filter bank. The modulation depth was calculated using the "with phase" method from chapter 5. For these analysis filter banks the modulation depth was sufficient to elicit temporal pitch cues in the three normal-hearing subjects (NH2, NH4 and NH5) and in all four CI subjects. For the GTF and BUT filter banks the modulation depth of the reference stimuli ranged from approximately 10% to 33% and from 11% to 36% respectively. This modulation depth was insufficient to generate discriminable temporal pitch percepts in any of the normal-hearing subjects while in contrast

it was sufficient for at least some stimuli for the three better CI subjects. Although the modulations for the GTF and BUT filter bank were undiscriminable in rate for the normal-hearing subjects, they were still detectable. The required modulation depth for modulation detection at 133 is in the order of 10% for normal-hearing subjects [Bacon and Viemeister, 1985]. The presence of these fluctuations may have elicited a roughness sensation [Zwicker and Fastl, 1999] that may have interfered with the place pitch cues and resulted in negative temporal pitch cues for the BUT filter bank and for some subjects with the GTF filter bank (see Figure 6.9).

This difference for rate discrimination as a function of modulation depth between the vocoder and the CI subjects may be overcome by the insertion of an additional compression/expansion stage into the vocoder. An expansion stage would increase the modulation depth and thus provide more effective temporal pitch cues. This suggested compression/expansion stage is not equivalent to the compression stage found in CI systems. The suggested compression/expansion stage is intended to overcome any residual differences in loudness growth (or more specifically perceptual modulation depth) between the CI with compression and the vocoder.

3. A third factor why the effect of temporal pitch cues was lower in the NH subjects using the vocoder compared to the effect in CI is the peripheral filtering of the normal ear. Section 6.4 and the study of Hanna [1992] showed that at low center frequencies the smaller bandwidth of the peripheral filters limits the effect of the temporal pitch cues. This effect is largest for the standard vocoder because for the standard vocoder the spectral region of the output signal is matched to the spectral region of the input signal and the original stimuli in the present study only contained energy in the lower frequencies because of the relatively low formant frequencies and the maxima selection in the processing. Consequently, the output signals of the standard vocoder in the present study only contained lower frequencies where the limiting effect of the peripheral filters exists. The F0-related modulations in the envelope of the different channels of the standard vocoder can be obscured by the modulations already present on the basilar membrane. Due to the limited bandwidth of the basilar membrane at lower frequencies the effective modulation depth of the F0-related modulations can be reduced.

   This effect is absent for the CISIM vocoder because the center frequency of the resynthesis filters are shifted up in frequency with respect to the center frequencies of the analysis filters. The stimuli processed with the CISIM vocoder contained energy at higher frequencies where the effect of the temporal pitch cues is not limited by the peripheral filters of the normal ear. There was a significant increase in the effect of temporal pitch cues between the standard and the CISIM vocoder (p = 0.031) for the filter bank with sufficient modulation depth (ACE) and for the three better performing NH subjects.

Summarizing the three previous factors, a noise band vocoder can be used as an acoustic model for temporal pitch research in cochlear implants 1) for normal-hearing subjects with relatively good pitch sensitivity, 2) for stimuli with enough

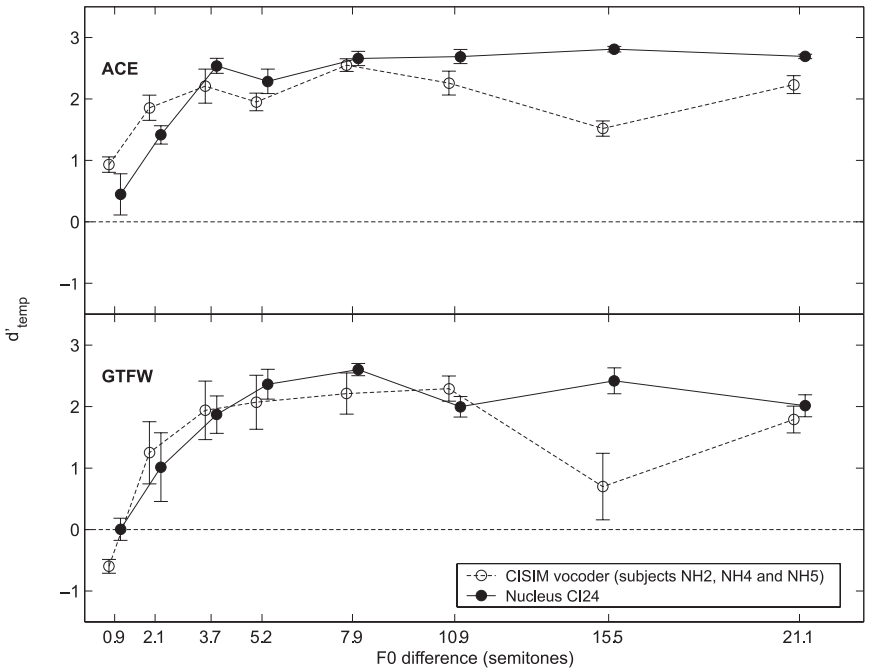


Figure 6.13: The amount of effect of the temporal pitch cues for F0-discrimination for stimuli processed with the CISIM vocoder for filter banks ACE and GTFW and as a function of relative F0 difference. The results are averaged over NH subjects NH2, NH4, and NH5 which have relatively good temporal pitch sensitivity. The results for four CI subjects from chapter 5 are included for comparison. For this reduced set of conditions and for these subjects, the CISIM vocoder succeeds in modeling the CI data. The error bars indicate ± 1 standard error of the mean of the inter subject variability.

modulation depth, 3) and for a vocoder with noise bands at higher frequencies (i.e. the CISIM vocoder). This is shown in Figure 6.13 where the average effect of the temporal pitch cues (d'_{temp}) for subjects NH2, NH4 and NH5 are depicted for the ACE and GTFW filter banks obtained with the CISIM vocoder. There exists relatively good correspondence between the results obtained with the CISIM vocoder for these subjects and for the results obtained with the four CI subjects from chapter 5.

Finally, a fourth factor may have contributed to the poor effectiveness of the temporal pitch cues for the NH listeners using the vocoder compared to the CI subjects. The noise-band carriers of the vocoder have inherent random envelope fluctuations, creating an "external" variability. These random modulations may interfere with the envelope modulations in the analysis-channel envelopes related to the F0 and eliciting the temporal pitch cues [Formby and Muir, 1988; Hanna, 1992]. Therefore, a deterministic carrier with limited envelope modulations in the temporal pitch frequency range might be more suitable as a carrier for a vocoder intended as

an acoustic model for pitch sensation in CI [Carlyon et al., 2002; van Wieringen et al., 2003; Deeks and Carlyon, 2004].

6.6 General discussion

Our results indicate that the relative contributions of spatial and temporal pitch cues for F0-discrimination can be altered by varying the width of the single-channel excitation patterns. Narrower excitation patterns with steeper slopes elicit highly salient place pitch cues and with very narrow excitation patterns these place pitch cues may become more salient than the temporal pitch cues.

In currently used CI-systems, F0-discrimination is mediated by temporal pitch cues more than by place pitch cues [Geurts and Wouters, 2001; Green et al., 2002; Chapter 5]. However, the present results suggest that narrower excitation patterns may provide more salient place pitch cues. Narrower excitation patterns were reported using bipolar or tripolar stimulation compared to the monopolar mode used in most current CI systems [Hartmann and Kral, 2004]. Ultimately, these narrow excitation patterns may elicit place pitch cues which are more salient than the temporal pitch cues. This might enhance F0-discrimination in CI subjects whereas the limit for the temporal pitch cues appears to be reached [Green et al., 2004].

In most studies using vocoders to investigate the effects of processing for pitch perception (or pitch related tasks) in CI subjects the carriers of the vocoders are spectrally matched, as in the standard vocoder in the present study. I.e., the analysis filters and the resynthesis filters are identical for noise-band vocoders [Fu et al., 1998; Faulkner et al., 2000; Green et al., 2002; Xu et al., 2002; Green et al., 2004; Kong et al., 2004; Fu et al., 2004; Qin and Oxenham, 2005] or pulse-train vocoders [Deeks and Carlyon, 2004], or the frequency of the sinusoidal carrier is set at the center frequency of the analysis filters for sinusoidal vocoders [Lan et al., 2004; Fu et al., 2004]. Firstly, our results indicate that the place (or spectral) pitch cues with the spectrally matched vocoders are more salient than those found in the CI subjects, especially for the case where analysis and resynthesis filters were completely identical as in the BUT condition for the standard vocoder. This suggests that the spectral pitch cues found in the mentioned vocoder studies may have been stronger than what may be obtained in CI subjects, especially when many channels are used in the vocoder. Secondly, our results also indicate that the temporal pitch cues obtained with the standard vocoder are less salient than the cues CI subjects can use because of the peripheral filtering in the normal ear. This suggests that with the spectrally matched vocoders of most other studies the obtained temporal pitch cues may be smaller than what may be found in CI subjects because the channels with center frequencies below 1 kHz are less effective in providing salient temporal pitch cues. Taking together the last two findings, it is possible that the spectral pitch cues were relatively more important than the temporal pitch cues with the vocoders than what may be obtained with CI subjects, especially as the number of channels would be high and the spectral resolution would be good. For sinusoidal vocoders the relative contribution of the spectral pitch cues may be even more exaggerated as the

excitation profile of sinusoids is very narrow and this leads to very discriminable spectral pitch cues.

Pitch or F0-discrimination solely based upon place pitch cues was strongly affected in the present study by spectral smearing (see Figure 6.11). This is in contrast with the smaller effect of spectral smearing on speech perception. Spectral smearing only affects speech understanding when the slopes of the noise-band carriers of the vocoder are more shallow than 18 dB/octave [Shannon et al., 1998]. This observation is consistent with the higher spectral resolution needed for melody recognition than for speech understanding [Smith et al., 2002].

There exists however a difference between the amount of spectral smearing necessary to replicate CI subjects' speech perception performance and the amount of spectral smearing necessary to replicate CI subjects' pitch discrimination performance. The mean speech-in-noise recognition thresholds of implant users are close to those of normal-hearing subjects listening to 4-channel spectrally smeared (with 6dB/octave resynthesis filters) noise-band vocoded speech [Fu and Nogaki, 2005]. In the present study the place pitch performance was matched for a 22-channel noise-band vocoder with slopes of 40 dB/octave (the CISIM vocoder). The cause of this difference in smearing necessary for matching CI performance is unknown, but it two factors may be important. Firstly, in the CISIM vocoder the spectrum is spectrally shifted and compressed along the cochlear axis. This probably affected performance because for a vocoder with a matched spectrum and shallower slopes (the standard vocoder in this study) the performance of place pitch discrimination was better even though the slopes of the carriers were shallower (24dB/octave). This is similar to the reduced speech understanding for spectrally shifted and compressed speech [Fu and Shannon, 1999; Baskent and Shannon, 2003]. In the studies assessing the effect of spectral smearing on speech perception the spectrum was not warped along the cochlear axis. Secondly, channel-interactions in CI-subjects may be greater during speech than during the stationary signals used in the present study because dynamic stimuli may cause stronger channel-interactions [Chatterjee, 2003].

Only post-lingually deafened CI subjects participated in the pitch discrimination experiments that were used to verify the acoustic model. However, both place pitch and temporal pitch mechanisms can be impaired in pre-lingually deafened subjects that were implanted at a relatively late age (approximately after 12 years of age) [Busby et al., 1993; Busby and Clark, 2000a; Busby and Clark, 2000b]. In those subjects often no pitch ordering across the electrode array is observed [Busby and Clark, 2000b]. Therefore, the acoustic model presented in this chapter is not likely to be generalized to this particular group of CI subjects. For pre-lingually deafened CI subjects that were implanted early in life, normal pitch perception can be observed [Busby and Clark, 2000a; Busby and Clark, 2000b]. In contrast to the previous group of subjects, these subjects experience auditory sensations in the 'critical period' of their brain development and this allows for at least partial maturation of their auditory system [Hartmann and Kral, 2004]. It may thus be possible to extend

the present acoustic model to also include early-implanted pre-lingually deafened CI subjects, although this remains to be verified in future experiments. The model thus appears applicable for post-lingually deafened CI subjects and early-implanted pre-lingually deafened subjects. These two groups constitute the major portion of all CI subjects.

The application of the presented acoustic model (the CISIM vocoder) to simulation of CI subjects in general may be complicated by the fact that the CI subjects participating in this study were all relatively good performers with very good electrode discrimination compared to other post-lingually deafened CI subjects [Nelson et al., 1995]. Secondly, as already mentioned, the performance on the pitch discrimination task of some of the NH subjects in this study was poorer than that of other NH subjects on similar tasks reported in other studies. Because of these two factors the pitch discrimination obtained with the acoustic model may be somewhat too optimistic to replicate CI performance in general. Especially the length constant λ could be considered somewhat longer, or equivalently more spectral smearing, in more general applications of the model or a compressive function might be used to limit the effectiveness of temporal pitch cues.

6.7 Conclusions

Although noise-band vocoders have been proven to provide a successful acoustic model to study speech perception in cochlear implants, it may not be straightforward to extend the model for pitch perception research in CI. The results of the present study indicated that both temporal and place pitch sensitivity can be affected by parameters of the noise-band vocoder.

1. The degree of spectral overlap between adjacent resynthesis filters of the noise-band vocoder is inversely proportional to the place pitch sensitivity subjects obtain with the vocoder.
2. The discrimination of noise bands filtered with the CISIM filter based on place pitch cues is independent of center frequency.
3. If the discrimination of adjacent channels with the noise-band vocoder, or equivalently single-channel place pitch sensitivity, is adjusted (by varying the spectral overlap of the channels) to match the electrode discrimination performance in CI subjects, then the place pitch sensitivity with the vocoder matches the place pitch sensitivity of CI subjects also for complex multi-channel stimuli.
4. Temporal pitch sensitivity is limited for lower frequency channels of the noise-band vocoder. For noise bands having lower center frequencies the peripheral auditory filters limit the modulation depth and consequently the temporal pitch sensitivity. A noise-band vocoder where the resynthesis filters all have high center frequencies (e.g. mimicking a shallow insertion depth) may overcome this effect of the peripheral auditory filters.

5. There is large variation in subjects' performance on pitch discrimination tasks, even for normal-hearing subjects.
6. A compression/expansion stage in the vocoder has to be adjusted to the effect of the compression in the CI-system. The resulting modulation depths in the neural excitation pattern with electrical stimulation and with acoustical stimulation may differ if the CI-compression is insufficient or too high. An extra compression or expansion function in the vocoder may compensate for these differences.

Taking these factors into account, a noise-band vocoder may be used as an acoustic model for pitch perception research in CI. Nevertheless, one has to take great care with the interpretation of the results because noise bands differ fundamentally from the deterministic pulse trains used in cochlear implants.

Chapter 7

Improved music perception with F0mod sound processing strategy

J. Laneau, M. Moonen, and J. Wouters. Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neuro-Otology,* In Review.

In the previous chapters it was shown that cochlear implant subjects can not tell the difference between two notes normal-hearing persons easily discriminate. In other words, pitch discrimination in CI subjects is very poor compared to NH performance. In chapter 5, we showed that with the current sound processing schemes for cochlear implants little improvement could be obtained. It was shown that pitch discrimination in CI subjects was still very poor whatever the filter bank or the envelope-detection mechanism.

In this chapter a new sound processing scheme (F0mod) is presented, based upon the results of the previous chapters, that is designed to improve pitch and music perception. The rationale behind the new sound processing strategy and the implementation are described in section 7.2. In the F0mod scheme, slowly varying

channel envelopes are explicitly modulated sinusoidally at the fundamental frequency of the input signal, with 100% modulation depth and in phase across channels to maximize temporal envelope pitch cues. Secondly, the filter bank of the F0mod scheme is designed to optimize place pitch cues based upon the results of chapter 5.

In the following sections, the performance of the new scheme for music and pitch perception is compared in four different experiments to that of the current clinically used sound processing scheme (ACE) in six Nucleus CI24 subjects. The results of the four experiments show that: 1) F0-discrimination of single formant stimuli was not significantly different for the two schemes (see section 7.3), 2) However, F0-discrimination of musical notes of five instruments was three times better with the F0mod scheme for F0 up to 250 Hz (see section 7.4), 3) Melody recognition of familiar Flemish songs (with all rhythm cues removed) was improved with the F0mod scheme (see section 7.5), 4) Estimates of musical pitch interval, obtained in a musically trained CI subject, matched more closely the presented intervals with the F0mod scheme (see section 7.6). These results indicate that explicit F0-modulation of the channel envelopes improves music perception in CI subjects.

7.1 Introduction

In contrast to the poor results obtained in CI subjects listening through their sound processors, a number of studies have shown that musical pitch perception is possible in CI subjects with synthetic stimuli where the pitch is varied by means of the pulse rate of a single channel pulse train [Pijl and Schwarz, 1995a; Pijl and Schwarz, 1995b; McDermott and McKay, 1997; Pijl, 1997]. With these synthetic signals subjects were able to recognize melodies without any rhythm cues and to correctly label musical intervals. Moreover, their performance approached the performance of naïve normal-hearing subjects. However, the sound processing strategies for CI, that were designed to optimize speech perception, failed to correctly encode pitch information so that musical pitch perception became impossible when harmonic acoustical stimuli were presented through the subjects' sound processors [Pijl, 1997].

In summary, current clinically used sound processors limit the performance of the subjects for music and pitch and variations in filter bank or envelope-detection mechanism to the current sound processors do no remove this limit (see chapter 5). To remove this limit, a number of recent studies have developed new sound processing schemes to improve pitch perception in CI subjects [Geurts and Wouters, 2001; Green et al., 2004; Lan et al., 2004; Geurts and Wouters, 2004].

Geurts and Wouters measured F0-discrimination for synthetic vowels in Laura CI subjects. In their first study [2001] they developed a new sound processing scheme based upon the CIS scheme that was designed to increase the modulation depth of the beatings in the channels related to F0. However, they found no significant difference with the standard CIS strategy. In the second study of Geurts and Wouters [2004], an alternative new sound processing strategy was developed that focussed on the place pitch cues available to the subjects. For this scheme, with increased resolution in the lower frequencies, Geurts and Wouters found improved F0-

discrimination based upon place pitch cues (and with temporal pitch cues removed). However, when the temporal pitch cues were added, they found little or no improvement with the new sound processing scheme compared to standard CIS. In chapter 5, we confirmed and extended this result and showed that it was very hard to combine salient place pitch cues and salient temporal pitch cues in the same sound processing scheme by merely adjusting the filter bank.

Lan *et al.* [2004] measured tone recognition and speech understanding for Mandarin Chinese in 20 normal-hearing subjects using a sinusoidal vocoder for 4, 6 or 8 channels. They compared standard CIS to a new scheme that extracted F0 from the input signal and used this information to frequency modulate the sinusoidal carriers of the vocoder. They found an improvement with the new scheme in all their test conditions. However, the envelopes of the comparison CIS strategy in their study were low-pass filtered at 50 Hz, and consequently all temporal pitch cues were removed from the stimuli in their standard CIS condition. Most probably, this deteriorated the performance of the subjects in the standard CIS condition significantly because temporal pitch cues are the most important cues for F0-discrimination with noise-band vocoders (see chapter 6) and in CI subjects (see chapter 5). Furthermore, their scheme has not been tested with CI subjects but only with a vocoder in normal-hearing subjects and it can be difficult to extrapolate pitch perception results obtained with vocoders to CI subjects (see chapter 6). Also the mechanical filtering in the normal ear elicits both place and temporal cues when a frequency modulated tone is presented. When the frequency modulation is only presented through a temporal code in a CI performance is very low and in a CI the place pitch cues would be a lot harder to present to the subject [Chen and Zeng, 2004].

Green *et al.* [2004] measured F0 glide identification (falling or rising) for synthetic diphthongs in normal-hearing subjects using a noise-band vocoder and in Clarion CI subjects. They compared the performance for a standard CIS strategy and for a new scheme where the envelope of each channel modulated a saw-tooth shaped pulse train instead of a steady-state pulse train in standard CIS. The saw-tooth waveform was periodic at the rate of the F0 and in phase over the different channels. They found a small but significant improvement in the glide labeling task with their new scheme(s). Because of the relatively small size of the improvement and possible technical problems with F0-extraction they concluded that the potential of their schemes in every day circumstances was limited. However, their methods were restricted to speech-like signals and musical pitch perception with the new scheme was not assessed.

In this chapter a new sound processing scheme is proposed that is designed to enhance (musical) pitch perception. The scheme is similar to the scheme proposed by Green *et al.* [2004] but incorporates an F0 extractor and the modulating waveform is different. The rationale and implementation of the new scheme are given in section 7.2. In the following sections the new sound processing scheme is tested perceptually and the performance of Nucleus CI24 subjects is compared to their performance with the current clinically used sound processing strategy for Nucleus (ACE). In the first two tests F0-discrimination is measured for stylized vowels (section 7.3) and for musical notes (section 7.4). In section 7.5 melody

recognition was measured using both sound processing schemes and in section 7.6, a musically trained CI subject performed a musical interval labeling task with notes processed through both sound processing strategies.

7.2 Sound processing scheme

7.2.1 Rationale of the F0mod scheme

The design of the new sound processing scheme fits into a general framework to optimize CI performance. The main goal of the framework is to present an auditory percept to the CI recipient that is minimally different from the sensation a normal-hearing subject would have listening to the same input sound. For example, a sound that would have a certain loudness for an average normal-hearing subject should be processed and delivered by the CI system to the CI subject such that the loudness for the CI subject is the same as for this average normal-hearing subject. An example of such a scheme has already been presented by McDermott *et al.*[2003] where the specific loudness of the electrical stimulus was designed to match the loudness according to a model for normal hearing. In the sound processing scheme proposed in this chapter, the focus of the percept is the pitch of an input sound. A computational model is used to extract the pitch normal-hearing subjects would perceive and the sound processing strategy is adjusted such that the resulting electrical pattern elicits an equally high pitch percept in the CI subjects.

The proposed sound processing scheme is designed to optimize pitch perception related to both periodicity pitch and spectral pitch. The periodicity pitch in normal-hearing subjects is presented to CI subjects by means of the CI temporal pitch cues, while the spectral pitch in normal-hearing subjects is presented by means of the CI place pitch cues.

7.2.1.1 Periodicity pitch

Periodicity pitch is derived from the peripheral neural excitation pattern by means of a spatio-temporal mechanism that is incompletely understood [Oxenham et al., 2004]. However, models based upon the autocorrelation function of the peripheral neural patterns can accurately predict the pitch perceived by normal-hearing subjects for most natural sounds occurring in the natural environment [Licklider, 1951; Meddis and Hewitt, 1991; Cariani and Delgutte, 1996]. The autocorrelation function can be regarded as an all-order statistic of the inter spike intervals, i.e. the autocorrelation function describes the intervals between a spike and all previous and following spikes.

In contrast, the temporal pitch percept in CI subjects is poorly modeled by an all-order statistic but is more closely related to a weighted sum of the first order spike intervals [Carlyon et al., 2002]. This means that only the timing between successive pulses determines the temporal pitch percept of CI subjects. Although this temporal pitch mechanism also exists in normal-hearing subjects (as the purely temporal pitch mechanism), it is not considered for these subjects in the present study, because for signals containing resolved harmonics, as do most natural sounds, the perceived pitch is dominated by the previously described periodicity pitch.

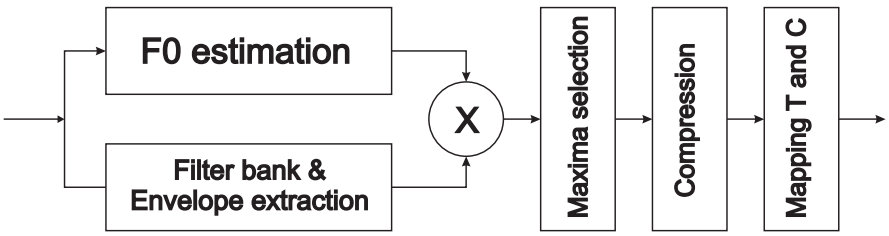


Figure 7.1: Overview of the processing blocks of the newly proposed sound processing scheme (F0mod). See text for details.

To account for the discrepancy in these pitch perception mechanisms, namely periodicity pitch in normal-hearing subjects and temporal pitch in CI subjects, the sound processing scheme presented here extracts the periodicity pitch using an autocorrelation based method and presents this extracted pitch through the first-order intervals of the modulation of the electrical pattern. Additionally, the presented stimulation pattern is designed to make the perceived temporal pitch as clear and as unambiguous as possible because all presented first-order intervals are constant across channels and evoked with maximal modulation depth.

7.2.1.2 Spectral pitch

Spectral pitch cues in NH subjects and place pitch cues in CI are both related to the centroid of the excitation pattern (see chapter 5 and [Zwicker and Fastl, 1999]). Consequently, the place pitch cues perceived by a CI subject will be matched to the spectral pitch cues as perceived by a normal-hearing subject when the excitation patterns along the basilar membrane are the same for electrical and acoustical stimuli. Unfortunately, in current CI systems the excitation pattern evoked by the electrical stimuli does not match the excitation pattern that is evoked in a normal-hearing subject presented with the same sound. This is because the center frequencies of the channels are not matched to the location along the basilar membrane of the corresponding electrodes, and consequently the excitation pattern is warped along the basilar membrane.

However, this issue can be solved for differences in place pitch between two stimuli when the warping of the excitation pattern is linear along the basilar membrane, because in this case the differences in place pitch perceived by the CI subject are proportional to the spectral pitch differences a normal-hearing subject perceives (see the Appendix in section 7.9). The filter bank of the sound processing scheme presented here is designed such that the electrically evoked excitation pattern is a linearly compressed version of the acoustically evoked excitation pattern in a normal-hearing person.

7.2.2 Implementation of the F0mod scheme

An overview of the processing blocks of the new sound processing scheme (F0mod) is shown in Figure 7.1. The input sound signal is presented to two parallel blocks. The first block, the filter bank, splits the signal into 22 channels and extracts the relatively slow varying envelopes of each channel. The second block estimates the

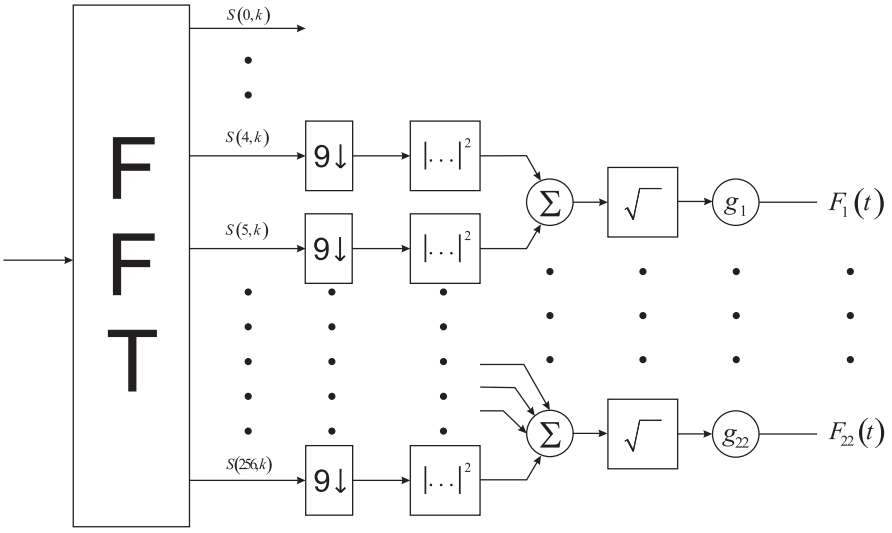


Figure 7.2: Block diagram of the processing stages in the filter bank of the F0-mod strategy. The input signal is analyzed using a 512-point FFT, then down sampled 9 times. The magnitude of the complex spectrum is calculated to obtain the Hilbert envelope. Subsequently, the different frequency bins are combined into 22 channels.

fundamental frequency of the input signal based on the autocorrelation of the input signal. The 22 output channels of the filter bank are then modulated sinusoidally at the estimated fundamental frequency. The next processing blocks perform maxima selection, compression of the waveform and map the waveform between the T and C levels of the subjects.

7.2.2.1 Filter bank

The filter bank of the new scheme is based upon the filter bank of the ACE strategy (see chapter 5). The main difference is the length of the FFT-buffer ($L = 512$ in the new scheme) which leads to higher frequency resolution and a lower cut-off frequency in the envelope extractor. The block diagram of the processing stages in the filter bank is given in Figure 7.2. The details of the implementation of the filter bank are discussed below.

The filter bank block reads in the input signal, sampled at 16 kHz, and stores it into buffers of 512 samples with 503 samples of overlap. This leads to an update rate F_{update} of approximately 1778 (=16000/9) buffers per second, equivalent to an effective 9-fold down sampling. The spectrum of each buffer is obtained through an FFT after applying a Hanning window using (5.2) and (5.3). The resulting spacing between adjacent frequency-bins of the spectrum is 31.25 Hz (=16000/512) and the bandwidth of the main lobe of each frequency-bin is 125 Hz [Oppenheim and Schafer, 1999]. The envelope is estimated, based upon the definition of the envelope in the Hilbert Transform, by calculating the magnitude of the complex spectrum. Because the bandwidth of each frequency-bin is 125 Hz, the extracted envelopes are

limited to 62.5 Hz. This cut-off frequency of the modulations is high enough to allow for good speech perception [Drullman et al., 1994; Shannon et al., 1995], but modulations at higher frequencies, leading to temporal pitch cues, are filtered out.

Because most speech signals do not contain energy below 100 Hz, the lowest 4 bins (including the DC bin) are discarded. Consequently, the filter bank spectrum ranges from 125 Hz up to 8000 Hz. The remaining bins are combined into 22 channels such that the bandwidths of these channels correspond to an equal number of ERB [Glasberg and Moore, 1990]. The assignment of the bins to each channel is shown in Table 7.1.

Table 7.1: Frequency allocation table for the 22 channels of the two filter banks that are compared in this study (F0mod and ACE). For the F0mod filter bank the channels are combinations of 257 bins from the 512-point FFT. For the ACE filter bank the channels are combinations of 65 bins from the 128-point FFT.

	F0mod			ACE		
Channel Number	Number of bins	Center freq. low bin (Hz)	Center freq. high bin (Hz)	Number of bins	Center freq. low bin (Hz)	Center freq. high bin (Hz)
1	2	125	156	1	250	250
2	2	188	219	1	375	375
3	2	250	281	1	500	500
4	3	313	375	1	625	625
5	3	406	469	1	750	750
6	3	500	563	1	875	875
7	4	594	688	1	1000	1000
8	5	719	844	1	1125	1125
9	5	875	1000	1	1250	1250
10	6	1031	1188	2	1375	1500
11	7	1219	1406	2	1625	1750
12	8	1438	1656	2	1875	2000
13	10	1688	1969	2	2125	2250
14	11	2000	2313	3	2375	2625
15	13	2344	2719	3	2750	3000
16	15	2750	3188	4	3125	3500
17	17	3219	3719	4	3625	4000
18	20	3750	4344	5	4125	4625
19	23	4375	5063	5	4750	5250
20	27	5094	5906	6	5375	6000
21	31	5938	6875	7	6125	6875
22	35	6906	8000	8	7000	8000

The amplitude of each channel is defined by the total RMS of the assigned bins:

$$F_n(t) = g_n \sqrt{\sum_q \left|S(q,t)\right|^2} \quad ,n = 1...22 \; and \; q = 0..256 \tag{7.1}$$

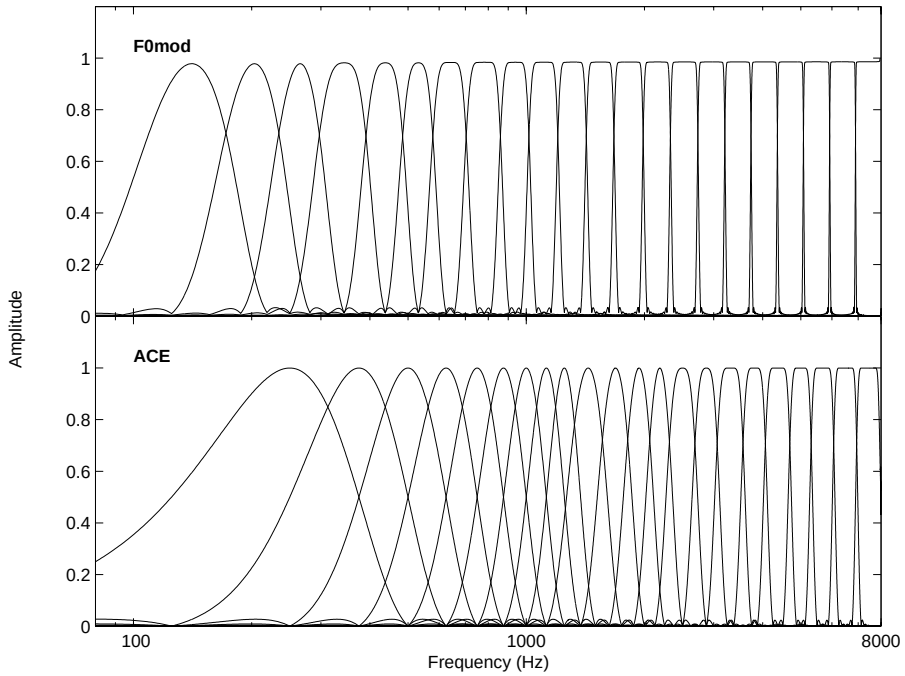


Figure 7.3: Frequency responses for the two filter banks used in the present study. The upper panel shows the frequency response of the 22 channels of the F0mod filter bank based upon a 512 point FFT. This filter bank is also used by the ACE512 strategy. The lower panel shows the frequency response of the 22 channels of the ACE strategy. The frequency response was obtained by measuring the output of each channel for sinusoids from 80 Hz to 8000 Hz.

Where $F_n(t)$ is the amplitude of filter bank output channel *n* at time instant *t*, and $S(q,t)$ is the q^{th} frequency-bin of the FFT of the buffer at time instant *t*. The weights g_n are applied as in the ACE filter bank (see chapter 5) and are determined by Eq. (5.5). The frequency response of the 22 channels of the filter bank is shown in the upper panel of Figure 7.3.

7.2.2.2 Fundamental frequency estimation and modulation

The second parallel block in Figure 7.1 estimates the fundamental frequency (F0) of the input signal. A single F0 estimate is obtained for each buffer processed in the filter bank. The fundamental frequency is estimated by taking the inverse of the lag corresponding to the maximum in the circular autocorrelation.

The circular autocorrelation $a_c(\tau,t)$ is calculated through the inverse Fourier transform (IFFT) of the power spectrum $|S(q,t)|^2$ obtained in the filter bank block to reduce computational complexity:

$$a_c(\tau,t) = IFFT\left\{ \left|S(q,t)\right|^2 \right\} = \frac{1}{L}\sum_{q=0}^{L-1}\left|S(q,t)\right|^2 e^{j\frac{2\pi}{L}q\tau} \quad \tau = 0...L \tag{7.2}$$

where τ indicates the lag in the autocorrelation and t is the time instant of the processed input buffer. The estimated fundamental frequency $F0_{est}(t)$ is given by:

$$F0_{est}(t) = \frac{F_s}{\tau_{\max}(t)} \quad with \quad a_c(\tau_{\max},t) = \max_{T\min<\tau<T\max}\left[a_c(\tau,t)\right] \tag{7.3}$$

where F_s is the audio sampling frequency (16 kHz), and $\tau_{\max}$ is the lag corresponding to the maximum in the circular autocorrelation function, between lags $T_{\min} = 27$ (1.7 ms) and $T_{\max} = 213$ (13.3 ms). The limits prevent the F0 detector to detect erroneous F0 values that are too high (above 593 Hz) or too low (below 75 Hz).

The described F0 estimator is relatively simple and may return incorrect F0 estimates for signals recorded in adverse circumstances, for instance with back ground noise. Many improved F0 estimators have been developed in the previous decades and continue to be developed [Markel, 1972; Rabiner et al., 1976; de Cheveigne and Kawahara, 2002; Nakatani and Irino, 2004; Lan et al., 2004]. However, the main goal of this study was not to investigate the effect of the F0 estimator but to assess the feasibility of an sound processing scheme that improves pitch perception in CI subjects. Also all stimuli used in the experiments were synthetic signals of which the F0 was, at least approximately, known a-priori. Therefore a simple F0 estimator could be used in this chapter.

Subsequently, the channel amplitudes obtained from the filter bank are modulated with a sine at $F0_{est}$ and with 100% modulation depth. All channels are modulated with the same function and in phase:

$$A_n(t) = \left[0.5 + 0.5\sin\left(\frac{2\pi}{F_{update}}\sum_{i=0}^{t} F0_{est}(i)\right)\right] F_n(t) \quad ,n = 1...22 \tag{7.4}$$

where $A_n(t)$ and $F_n(t)$ are the modulated amplitude of channel *n*, and the output of the filter bank, respectively, at the time instant *t*. F_{update} is the update rate of the filter bank (1778 Hz).

The length of the buffers, in the filter bank and F0-estimator, was chosen so that the highest modulation frequencies of the filter bank outputs $F_n(t)$ were limited to 62 Hz (=1/2*Bandwidth of frequency bin; see above), and fundamental frequencies as low as 75 Hz can be estimated. Because the circular autocorrelation is symmetric, only 256 lags (=L/2) are present, and consequently the lowest detectable F0 is 62.5 Hz ($= F_s/(L/2)$). Because the modulations of the filter bank in the F0mod strategy only contained frequencies up to 62 Hz, there is no interference with the modulations imposed by the F0 estimator.

The proposed sound processing strategy (F0mod) imposes the maximum modulation depth on all channels and makes this modulation in phase. In chapter 5 we showed that F0-discrimination improves when the modulation depth of the F0 related modulations in the channel amplitudes increases and that this increase is clearest when the modulation is in phase over the different channels. Consequently, the imposed modulations make the temporal pitch cues as clear as possible (maximal modulation depth) and as unambiguous as possible because only a single first-order interval is present in the modulation across channels.

7.2.2.3 Further processing

The following processing blocks of the sound processing scheme are shown on the right hand side of Figure 7.1. In the first processing block (*Maxima selection*), only the amplitudes of the 8 channels with the largest amplitude are retained per processed buffer and the other channels are set to zero (see Eq. (5.16)). Subsequently in the *Compression* block, all channel amplitudes are compressed to accommodate for the reduced dynamic range of CI subjects and the steep loudness growth with increasing electrical current. The compression function is identical to the compression function used in the clinical Cochlear SPrint speech processor and is described in Eq. (5.17) of chapter 5. Finally, the resulting amplitudes are linearly mapped between the T and C levels for the appropriate channels of each subject (see Eq. (5.18)). The resulting current amplitudes $cl_n(t)$ are expressed in clinical current units which have an exponential scale (see Eq. (5.19)). The T and C-levels for each channel were clinically obtained using stationary pulse trains of 1800 pulses per second.

The obtained current amplitudes were used to modulate pulse trains of 1800 pulses per second per channel. Consequently the total stimulation rate was 14400 pulses per second. The pulses on different channels were presented interleaved in time and ordered from base to apex. The pulses were biphasic pulses with a phase duration of 25 μs and an inter phase gap of 8 μs. The stimulation mode was monopolar with both return electrodes active (MP1+2).

To accommodate for the small difference between the update rate of the filter bank (1778 Hz) and the stimulation rate per channel (1800 Hz), the channel amplitudes are resampled by duplicating every 81st sample. This equalizes the sampling rate of the channel amplitudes with the stimulation rate.

7.2.3 Comparison strategies

The sound strategy presented in the previous section (F0mod) is compared in the present chapter to two other strategies. The first strategy is the ACE strategy as it is implemented in Cochlear SPrint processors. The ACE strategy is defined in chapter 5 (section 5.2.3.1). The second strategy is a variation to the F0mod strategy, with the modulation of the envelopes discarded so that temporal pitch cues are not present (denoted ACE512 in this chapter).

The ACE and ACE512 strategies are similar to the processing scheme presented in Figure 7.1 but do not contain a F0-estimation block and the filter bank output

channels are not explicitly modulated. Or for both ACE and ACE512 this can be expressed as:

$$A_n(t) = F_n(t) \quad n = 1...22 . \tag{7.5}$$

1. The ACE strategy is defined in section 5.2.3.1 and the reader is refered to this section for details. The ACE filter bank reads in the input signal in buffers of 128 samples with 119 samples of overlap, leading to an update rate of 1778 Hz as in the F0mod strategy. (This differs from the implementation of the ACE strategy in chapter 5, where the update rate was 16000 Hz, and the output was resampled later. In the current implementation the down-sampling was moved forward in the block-processing scheme to optimize computational complexity. Also note that the down-sampling in the current implementation was only 9-fold, and 18-fold in the implementation of chapter 5.) The buffers are analyzed using a 128-point FFT after application of a Hanning window. The frequency-bin spacing is 125 Hz and the bandwidth of the main lobe of each frequency-bin is 500 Hz. The envelope is extracted by calculating the magnitude of the spectrum. The lowest two bins are discarded and the remaining bins are combined into 22 channels using Eqs. (5.4) and (5.5). The assignment of the bins to their respective channels is given in Table 7.1. The relatively wide bandwidth of the bins in the ACE filter bank allows having frequencies in the extracted envelopes up to 250 Hz. This includes modulation frequencies that elicit temporal pitch cues, without explicitly imposing them as in the F0mod strategy. The further processing of the ACE strategy is identical to the further processing of the F0mod strategy.

2. Because the ACE strategy and the F0mod strategy differ both in the filter bank used and in the presence/absence of the explicit modulation, a third strategy is added that only differs from the ACE strategy with regard to the filter bank and only differs from the F0mod strategy in the absence of explicit modulation. This ACE512 strategy has the same filter bank as the F0mod strategy, based upon a 512 point FFT. However, in contrast to the F0mod strategy but similar to the ACE strategy, the extracted envelopes are not explicitly modulated at F0. This means that only modulations up to 62 Hz are present in the output channels and consequently only place pitch cues can be elicited by stimuli processed with the ACE512 strategy.

7.3 Experiment 1: F0-discrimination of single-formant stimuli

7.3.1 Methods

7.3.1.1 Subjects

Six experienced post-lingually deafened CI recipients implanted with a Nucleus CI24 implant participated in the present study. The subjects were paid for their collaboration. Some relevant details about the subjects can be found in Table 7.2.

All subjects are relatively good performers with their implant in daily life. Subject CI5 used the LAURA implant before being reimplanted with a Nucleus CI24 implant 11 months before the experiments begun. All subjects except subject CI6 had previous experience in psychophysical experiments with pitch perception. Subject CI6 was a professional musician before becoming deaf. Subjects CI1, CI2, CI3, and CI4 are the same subjects that participated in the experiments of chapters 3 and 5.

Table 7.2: Relevant information about the subjects who participated in the experiments

Subject	Duration of profound deafness (y)	Etiology	Implant Type	Speech Processor	Clinical speech processing strategy	Age (y)	Implant experience at start of experiments (y;m)
CI1	11	Hereditary	Nucleus CI24R(ST)	ESPrit	Speak	20	3;1
CI2	3	Unknown	Nucleus CI24R(ST)	ESPrit	Speak	23	4;5
CI3	30	Progressive	Nucleus CI24R(CS)	SPrint	ACE	56	2;0
CI4	32	Progressive	Nucleus CI24R(CS)	ESPrit 3G	ACE	50	1;8
CI5	3	Menière	Nucleus CI24R(CS)	SPrint	ACE	47	7;5
CI6	30	Usher	Nucleus CI24R(CS)	ESPrit 3G	ACE	65	0;4

7.3.1.2 Stimuli

Fundamental frequency discrimination was measured for stylized vowels consisting of single formant stimuli. The stimuli were a subset of the stimuli used in chapter 5. The stimuli were generated by filtering a 500 ms pulse train sampled at 16 kHz with a cascade of two filters. The first filter was a second order IIR low-pass filter with a cut-off frequency of 50 Hz so that the output of the first filter resembled the glottal volume velocity. The second filter was a second order IIR resonator that created a formant in the spectrum. The details of the filters can be found in section 5.2.3. The formant frequency was 300 Hz, 350 Hz, 400 Hz, 450 Hz or 500 Hz and the bandwidth of the formant was kept fixed at 100 Hz.

The fundamental frequency of the reference signal was 133 Hz and the fundamental frequencies of the comparison signals were 140 Hz, 150 Hz, or 165 Hz (corresponding to relative F0 differences of 0.9, 2.1, or 3.7 semitones respectively). All stimuli were set to equal RMS power and then maximally amplified with a common gain while the signal was kept between -1 and 1 so that most of the stimuli optimally used the dynamic range of the compression block in the sound processing. The resulting RMS of the signals before processing was approximately 40% of the maximum input amplitude. All stimuli were processed offline with the three sound processing strategies (F0mod, ACE and ACE512), and subsequently stored on disk.

7.3.1.3 Psychophysical procedure

Fundamental frequency discrimination was measured in a 2 Interval - 2 Alternative Forced Choice (2I-2AFC) paradigm similar to the procedure used in chapter 5. The subjects were presented with the reference stimulus and the comparison stimulus in random order and were asked to indicate the highest one in pitch. The F0 of the reference stimulus was 133 Hz and the F0 of the comparison stimulus was set at one of the three higher F0 frequencies but within one trial both the reference and the comparison stimulus had the same formant frequency. No feedback was presented to enforce subjects to use their intuitive sense of pitch and to prevent the subjects from using cues other than pitch (for instance loudness). Reference and comparison stimulus were separated by a 500 ms silent gap.

The trials were presented in separate blocks for each of the three sound processing schemes. Each block contained trials with all three comparison frequencies for all 5 formant frequencies, and within each block each trial was repeated four times. Different formant frequencies were included per block in order to prevent subjects to identify and learn the reference stimulus based upon other cues. In total 3 different blocks (for each strategy) consisting of 60 trials were presented to every subject. The order of the blocks was randomized and each block was repeated twice.

To reduce the influence of loudness cues, amplitude was roved by multiplying the amplitude of all stimuli upon each presentation with a random factor between 0.77 and 1, with amplitude expressed in clinical current units and relative to the threshold level. The stimuli were presented and responses were collected using the APEX research platform (see chapter 2).

7.3.2 Results

The average proportion of correctly ranked trials is shown in Figure 7.4 for the three different sound processing schemes and as a function of frequency difference. The proportions were averaged over the five formant frequencies and over the six subjects.

A repeated measures analysis of variance (ANOVA) was performed on the obtained proportions with as within-subjects variables the F0-difference, sound processing scheme, formant frequency, and the measurement run (first or second presentation of the block), using the lower-bound statistics. The results showed a significant effect of both sound processing scheme [$F(1) = 75.9$; $p < 0.001$] and relative F0-difference [$F(1) = 18.3$; $p = 0.008$]. There was also a significant interaction effect between sound processing scheme and relative F0-difference [$F(1) = 10.2$, $p = 0.024$]. No other within-subjects factor or interaction of within-subjects factors had a significant effect. There were significant between-subjects differences [$F(1) = 1386.8$; $p < 0.001$].

Pairwise comparisons using the Least Significant Difference (LSD) showed that the subjects performed significantly worse with the ACE512 scheme than with the ACE and F0mod schemes ($p < 0.001$ for both schemes). But there was no difference between the results obtained with the ACE and F0mod schemes ($p = 0.470$). This lack of difference was caused by the difference between subject CI4 and the other

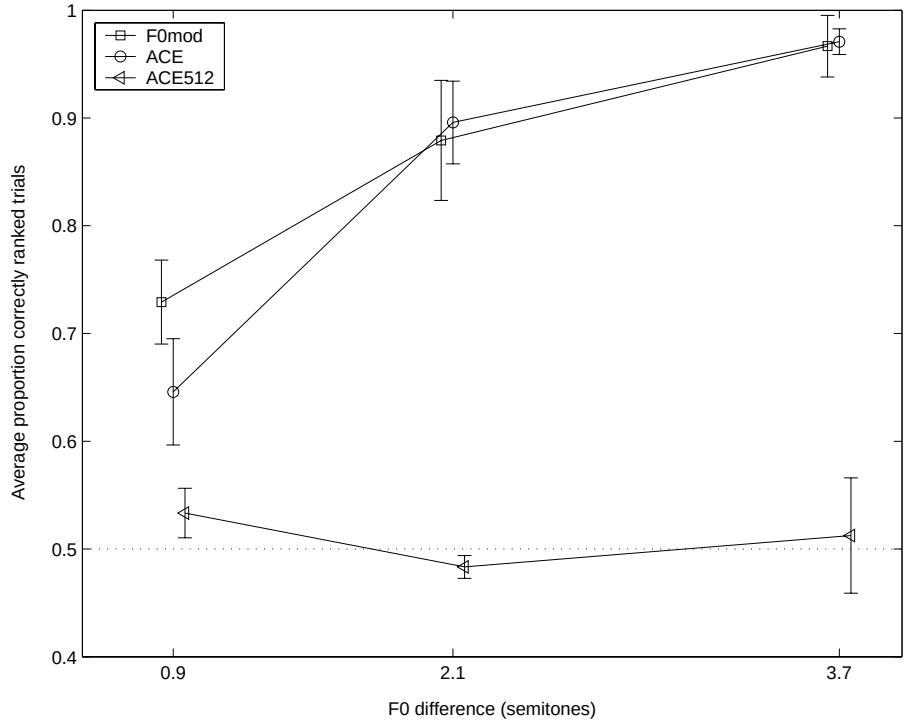


Figure 7.4: Results for the F0-discrimination of single formant stimuli with a reference frequency of 133 Hz. Each line shows the proportion of correctly ranked trials averaged over the six subjects as a function of relative F0 difference for the three different processing schemes. The dotted line indicates chance level. The error bars indicate ± 1 standard error of the mean over the six subjects.

five subjects. Subject CI4 obtained a lower F0-discrimination performance with the F0mod scheme compared to the ACE scheme. In contrast, all other subjects obtained better F0-discrimination performance with the F0mod scheme compared to the ACE scheme. When subject CI4 was dropped from the analysis a significant improvement was found for the F0mod scheme with respect to the ACE scheme ($p = 0.005$).

7.3.3 Discussion

With the ACE512 scheme subjects were unable to discriminate the F0 differences because all proportions are close to chance level. Increasing the relative F0 difference up to almost four semitones did not result in an improved F0-discrimination performance. This emphasizes the importance of temporal pitch cues for F0-discrimination in this condition because with the ACE512 no temporal pitch cues were present in the stimuli (while retaining the slow frequency modulations important for speech understanding [Drullman et al., 1994; Shannon et al., 1995]). However, the ACE and F0mod schemes are able to provide temporal pitch cues in the stimuli, and F0-discrimination reached a proportion of 75% correct between 1 and 2 semitones in the CI subjects.

This is consistent with the results of previous studies emphasizing the importance of temporal pitch cues for stimuli with fundamental frequencies below 300 Hz (see chapter 5).

The fact that no significant difference is found between the F0mod scheme and the ACE scheme is consistent with the results of Geurts and Wouters [2001]. Geurts and Wouters developed a signal processing scheme (F0 CIS) designed to enhance the modulation depth of the temporal fluctuations in the processed channels related to the fundamental frequency. They found no difference between this enhanced strategy and a standard CIS strategy where the fluctuations in the extracted envelopes resulted from interaction between different harmonics falling into the same filter bank channel, similar to the ACE scheme in the present study.

When the results of subject CI4 are excluded, the results of this study also agree with the results of Green *et al.* [2004]. Green *et al.* measured discrimination of direction of F0-glides in diphthongs for a standard CIS strategy and a modified strategy where the authors explicitly modulated all channels at the F0, similar to the F0mod scheme in the present study. Green *et al.* used a saw-tooth waveform or a 'sharpened' saw-tooth waveform to modulate the channel. The channels in the F0mod scheme of the present study were modulated sinusoidally. The results of Green *et al.* showed a significant improvement with the sharpened saw-tooth modulating waveform over the standard CIS strategy. However Green *et al.* also noted that this difference was relatively small. The average slope of logistic regressions fitted to the results of Green *et al.* was approximately 1.3 times higher with the 'sharpened' saw-tooth modulating waveform for the lower center frequency tested (113 Hz). The average slope of cumulative normal distributions fitted to the proportion of correctly ranked trials was 1.35 times higher with the F0mod strategy compared to the ACE strategy when averaged over all subjects except CI4. Thus, the improvements found in the study of Green *et al.* and in the present experiment for all subjects except CI4 are very similar.

7.4 Experiment 2: F0-discrimination for musical notes

7.4.1 Methods

F0-discrimination was measured for notes of five different instruments and at three different reference fundamental frequencies. The musical notes were generated through a software MIDI generator based on the Creative Labs Live sound font wave table. The selected instruments were grand piano, clarinet, trumpet, guitar, and synthetic voice. The midi notes, dotted eighth notes played at 120 beats per minute, were recorded at 16 kHz with 16 bit resolution. Consequently, the sustained portion of each note was approximately 400 ms. All recordings were truncated to 500 ms and set to equal RMS. All stimuli were processed with the three sound processing schemes offline and stored on disk.

F0-discrimination was measured for three reference F0's: 130.8 Hz, 185.0 Hz, and 370.0 Hz (corresponding to the notes C3, F3# and F4#, respectively). The difference

between the first two reference frequencies is six semitones or half an octave. The difference between the last two reference frequencies is exactly one octave. The F0 of the comparison signals were 1, 2 or 4 semitones higher (corresponding to approximately 6%, 12% or 26% relative F0-difference).

The same subjects participated and the same procedures were used as in experiment 1 of this chapter (section 7.3). Upon each trial the subjects were presented with a reference note (at one of the reference frequencies) and a comparison note (with higher F0) in random order. The subjects were asked to indicate the highest one in pitch. Both notes in a particular trial were played by the same instrument. The trials were presented in separate blocks for each combination of sound processing scheme and reference F0. Within each block the three comparison F0-differences were presented for all five instruments and every trial was repeated four times. Consequently, this lead to 9 different blocks consisting of 60 trials that were presented to each subject twice. Loudness roving was applied as in experiment 1 of this chapter.

7.4.2 Results

The average proportions of correctly ranked trials are shown in Figure 7.5 for each sound processing scheme and as a function of relative F0 difference. Each panel shows the results for a different reference F0.

A repeated measures ANOVA was performed on the proportions with five within-subjects factors: sound processing scheme (F0mod, ACE, or ACE512), reference F0, musical instrument or timbre, relative F0 difference, and the measurement run (first or second presentation of the block) using the lower-bound adjustment for the degrees of freedom. The analysis indicated significant effects of the sound processing scheme [$F(1) = 9.87$; $p = 0.026$], and the relative F0 difference [$F(1) = 76.0$; $p < 0.001$]. There were also significant interaction effects between sound processing scheme and relative F0 difference [$F(1) = 8.12$; $p = 0.036$] and between the reference F0 and relative F0 difference [$F(1) = 7.05$; $p = 0.045$]. The interaction effect between sound processing scheme and reference F0 approached significance [$F(1) = 5.82$; $p = 0.061$]. The subjects differed significantly in their performances [$F(1) = 10361$; $p < 0.001$].

The largest difference between the three sound processing schemes is found at the lowest reference frequency (130.8 Hz), shown in the left panel of Figure 7.5. Pairwise comparisons showed that for this reference F0 the ACE512 scheme produced lower scores than the ACE scheme ($p < 0.001$) and that the ACE scheme produced lower scores than the F0mod scheme ($p < 0.001$). The ACE512 scheme, which only provides place pitch cues, had the poorest performance. However, the ACE and F0mod scheme, which provide both temporal and place pitch cues, enabled the subjects to discriminate F0 differences of 2 to 4 semitones.

At the middle reference F0 (185.0 Hz) a pairwise comparison indicated no difference between the ACE and ACE512 schemes ($p = 0.797$). However, the results with the F0mod scheme were significantly higher than with the other two schemes ($p < 0.001$). Only with the F0mod scheme the average proportion of correctly ranked trials was above 75%, already for relative F0 differences of 2 semitones.

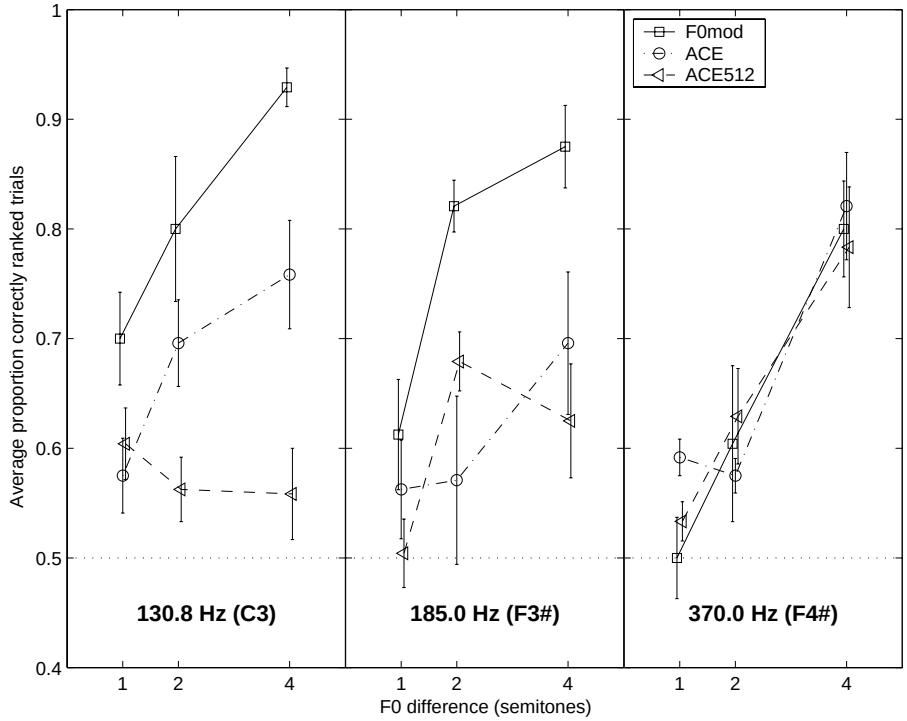


Figure 7.5: F0-discrimination results for musical notes of five different instruments. The results are averaged over the instruments and the six subjects. Each line shows the proportion of correctly ranked trials as a function of relative F0 difference for the three processing schemes. The results for the reference frequency fixed at 130.8 Hz, 185.0 Hz, and 370.0 Hz are shown in the left, middle and right hand side panel respectively. The dotted line indicates chance level. The error bars indicate ± 1 standard error of the mean over the six subjects and five instruments.

At the highest reference F0 (370.0 Hz) pairwise comparisons indicated no differences between any of the sound processing schemes. In this condition, all three sound processing schemes provided approximately the same level of performance to the subjects. On average, the proportion of correctly ranked trials increased monotonically with relative F0 difference and reached 75% at around 4 semitones.

7.4.3 Discussion

The F0mod scheme gave the best performance in the two lower reference F0 and did not differ from the other sound processing schemes at the highest reference F0. The average slope of a normal cumulative distribution fit to the average proportions was 2.6 and 3.3 times higher for the F0mod strategy than for the ACE strategy for the reference F0 of 130.8 Hz and 185.0 Hz, respectively. This means that the just noticeable F0 differences are approximately 3 times smaller with the F0mod scheme than with the ACE scheme in these conditions.

For both sound processing schemes providing temporal pitch cues (F0mod and ACE), the performance drops when the reference F0 frequency increases. This decrease in performance with increasing F0 was also found in chapter 5 and in the study of Green *et al.* [2004]. This decrease is most probably due to the reduced effectiveness of temporal pitch cues at higher modulation rates [Shannon, 1983a].

For the ACE scheme there may also have been an additional factor causing the decrease in F0-discrimination performance with increasing F0. Although the width of the main lobe of the bins in the ACE filter bank was 512 Hz, the transition width of the Equivalent Kaiser window is only 313.2 Hz [Oppenheim and Schafer, 1999]. This means that for higher F0 the additional harmonics in a filter bank bin are attenuated. Consequently, for F0 above 156.6 Hz the modulation depth is reduced. We showed in chapter 5 that deeper modulation depths may allow better F0-discrimination.

At the highest reference F0 tested (370 Hz), the subjects did not use any temporal pitch cues. For the ACE and ACE512 schemes there were no temporal pitch cues present in the stimuli because the F0 is above the highest modulation frequency present in the channels of those schemes (250 Hz, and 62.5 Hz, respectively). Furthermore, the subjects were unable to make use of the temporal pitch cues present in the stimuli processed with the F0mod scheme, because there was also no difference between the ACE512 scheme and the F0mod scheme.

Identical results were obtained with the 128-point FFT filter bank (ACE) and the 512-point FFT filter bank (ACE512 and F0mod) at the highest reference F0 tested, although the latter filter bank has more resolution in the lower frequencies. This is in contradiction with the results of chapter 5, where we found a benefit for the place pitch cues with filter banks having more resolution in the lower frequencies. However, in that study the difference in resolution in the F0 frequencies between the filter banks was larger than the difference in resolution between the filter banks of the present study. Moreover, in the experiments in chapter 5 the fundamental frequency component had a large effect on the overall spectral shape of the stimuli because the formant frequencies were relatively low (between 300 Hz and 500 Hz). In the present experiment the 'weight' of the fundamental on the overall spectral shape (or spectral centroid) was less prominent because the stimuli contained more harmonics.

The large difference found between the F0mod scheme and the ACE scheme for the lower reference F0, is in contrast with the results of experiment 1 of this chapter and with the results of Geurts and Wouters [2001] and Green *et al.* [2004]. In these experiments no or only a minor improvement was obtained with the sound processing scheme providing enhanced temporal modulations. However, all these studies used synthetic vowels or synthetic diphthongs generated based upon the Klatt speech synthesizer [Klatt, 1980]. In this synthesizer the stimuli are generated by filtering periodic pulse trains, a harmonic signal with all components of equal amplitude and in cosine phase. Harmonics summed in cosine phase and of approximately equal amplitude are known to generate a very clear envelope modulation. However, the envelope modulation of real-life signals can be 'smeared' when the signal is affected by reverberation or when noise is added to the signal.

Qin and Oxenham [2005] measured F0-discrimination with a noise-band vocoder to simulate CI-performance for various amounts of reverberation. They found the best performance for signals with all harmonics in cosine phase. However, when the harmonics were not in cosine phase the performance dropped significantly for vocoders with up to 8 channels. Consequently, the good performance of the ACE scheme in experiment 1 and the standard CIS schemes in Geurts and Wouters [2001] and Green *et al.* [2004] was probably due to the artificial nature of the stimuli in anechoic conditions. In this condition, the envelope modulations in the channels were already close to optimal to provide temporal pitch cues with the ACE scheme. Further enhancing the temporal pitch cues then did not result in large improvements.

However, with more realistic signals, as with the stimuli in the present experiment, the performance drops for the standard processing schemes as the modulation depth is reduced because of the temporal smearing. Consequently, the results of the subjects with the ACE scheme are better with the single formant stimuli of experiment 1 than with the musical notes of experiment 2. This difference in performance is much smaller with the F0mod scheme because the modulation depth is preserved in all cases.

7.5 Experiment 3: Melody recognition

7.5.1 Methods

Although an improvement was found for F0-discrimination with the F0mod scheme in the first two experiments of the present study, it can be argued that it is questionable that this improvement will lead to an improvement in music perception with CI subjects. In the present experiment we therefore compare melody recognition of familiar melodies with the ACE and the F0mod strategy in four CI subjects.

7.5.1.1 Stimuli

The most characteristic parts of 19 popular Flemish melodies (mostly nursery songs) were composed in MIDI format. The melodies were selected for their general familiarity and simplicity in rhythm. Subjects also use rhythmic cues to recognize melodies apart from pitch cues and this is especially so for CI subjects. However, in the present experiment we are only interested in the effect of the pitch cues. Therefore, the rhythms of all melodies were adjusted so that all melodies contained 16 quarter notes (by splitting longer notes into multiple quarter notes and by lengthening or dropping shorter notes). The notes of each melody were then transposed such that the median note of each melody was equal to F4# (370.0 Hz). The notes of the melodies are shown in Table 7.3.

The MIDI-melodies were rendered with the Clarinet instrument at 120 beats per minute using the software synthesizer described in experiment 2. Each melody was rendered twice: in a high register (around F4#; 370.0 Hz) and in a low register (around F3#; 185.0 Hz) for which all the notes were transposed one octave down. The recordings were made at 16 kHz with 16 bit resolution and truncated at a

duration of 8.15 s. All stimuli were then processed with the ACE and F0mod sound processing schemes and stored on disk.

Table 7.3: MIDI note values for the 19 melodies used in the melody recognition test. MIDI note value 60 corresponds to C4 (261.6 Hz) and each step denotes a single semitone.

Name of Song	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Broeder Jacob	62	64	66	62	62	64	66	62	66	67	69	69	66	67	69	69
Drie Koningen	66	66	64	63	66	66	64	63	63	66	66	68	68	66	66	66
Er is een Kindje geboren op aard	62	66	66	64	62	62	66	69	69	71	69	69	62	66	66	64
Handjes Draaien	62	66	69	66	67	64	71	69	62	66	69	66	69	69	62	62
Happy Birthday	62	64	62	67	66	66	62	64	62	69	67	67	62	74	71	67
In het bos	57	66	66	66	57	66	66	66	57	66	66	71	69	67	67	67
Klappen in de handjes	59	61	63	64	66	66	66	66	68	68	68	68	66	66	66	66
Altijd is Kortjakje	61	61	68	68	70	70	68	68	66	66	65	65	63	63	61	61
Lang zal hij leven	63	63	63	58	67	67	67	63	70	70	72	68	65	65	65	70
Old MacDonald	66	66	66	61	63	63	61	61	70	70	68	68	66	66	66	66
Schipper Mag Ik	66	66	66	68	66	63	59	63	66	66	68	68	66	66	66	66
Sinterklaas Kapoentje	66	66	68	68	66	66	63	63	66	66	68	68	66	66	63	63
In een klein Stationnetje	65	65	65	67	69	69	69	69	67	65	67	69	65	65	60	60
Zie ginds komt de Stoomboot	61	66	66	70	68	68	71	65	65	68	66	66	61	66	66	70
Sur le pont d'Avignon	65	65	65	65	67	67	67	67	69	70	72	65	64	65	67	60
k'Zag Twee Beren	69	69	69	74	66	66	66	69	62	62	62	66	69	69	64	64
Vreugde Symfonie	66	66	67	69	69	67	66	64	62	62	64	66	66	66	64	64
Vrolijke, Vrolijke Vrienden	64	59	64	66	59	66	68	68	66	64	64	64	69	69	69	68
Zakdoek Leggen	66	68	66	63	66	68	66	63	66	66	68	68	66	66	63	63

7.5.1.2 Procedure

Subjects CI1, CI3, CI4, and CI5 participated in the present experiment. Each subject was presented with the names of the 19 melodies and was asked to indicate their familiarity with each song. For each subject the 10 most familiar songs were selected. Each subject was presented with blocks of melodies presented in random order and had to indicate the name of the melody from a closed set. Each melody was presented twice within each block, so that each block consisted of 20 melodies. Four different blocks were presented to the subjects for each combination of sound processing scheme (ACE or F0mod) and register (High or low) and the different blocks were presented in random order. Each block was presented five times to each subject spread over two or three test sessions. No feedback was given to the subjects.

In the beginning of each test session the subjects were presented with each of their 10 most familiar melodies processed with the ACE scheme and with the original correct rhythm as an aid for the subjects to remember the melodies.

7.5.2 Results

Figure 7.6 shows the results for the melody recognition task averaged over the 5 runs and for each subject separately. The number of correctly identified melodies per block was analyzed using a repeated measures ANOVA with three factors: sound processing scheme, register of the melodies (high or low), and measurement run number (one through five). There was a significant difference between the sound processing schemes [$F(1) = 100.1$; $p = 0.002$] but there was no difference between

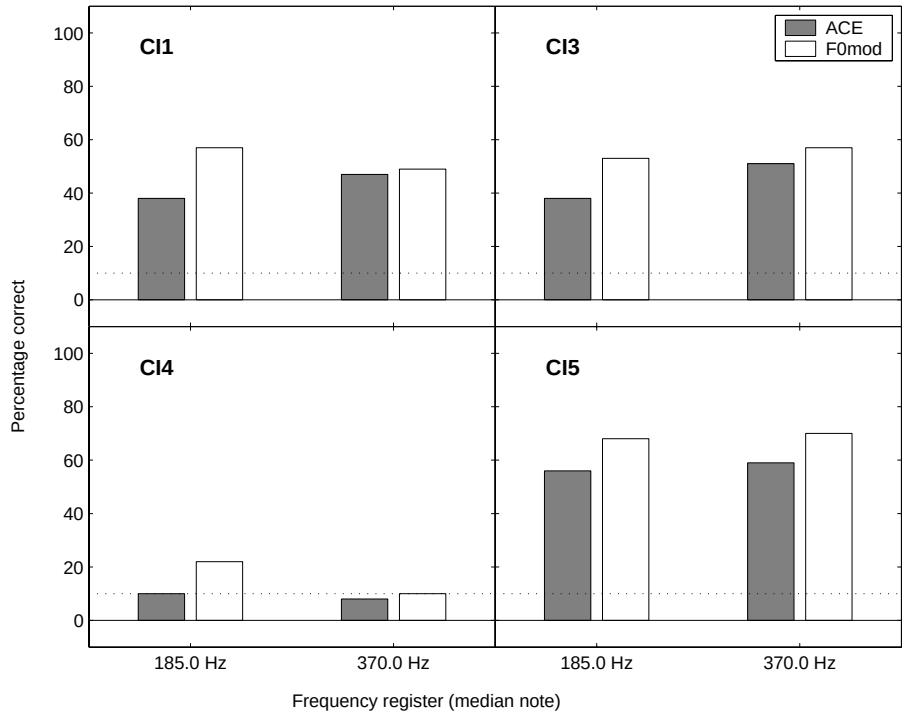


Figure 7.6: Melody recognition of familiar songs without rhythmic cues. Each panel shows the results for a specific subject. The dotted line indicates chance level. Each bar is the result of 100 trials. The subjects scored significantly (at 0.05 level) above chance when the proportion was above 15%.

the registers [F(1) = 0.124; p = 0.748]. There was also a slightly significant effect of the measurement run [F(1) = 10.343; p = 0.049]. There were no significant interaction terms. The subjects differed significantly in their performance [F(1) = 15.98; p = 0.028].

The subjects were able to recognize more melodies with the F0mod scheme than with the ACE scheme in both the high and the low frequency register. Subjects CI1, CI3, and CI5 scored relatively well on the melody recognition task with percentages of melody recognition ranging from 38% up to 70% correct. Subject CI4, scored a lot worse and reached only significantly above chance level for the F0mod scheme in the lower register.

Subjects CI, CI3, and CI5 also exhibited a small training effect, as the number of correctly identified melodies increased monotonically for all conditions with the measurement run number. The average percentage of correctly identified melodies was approximately 25% higher in the fifth experimental run compared to the first. This result is remarkable as no feedback was presented to the subjects.

7.5.3 Discussion

All subjects scored better with the new F0mod scheme than with the standard ACE strategy, at least for melodies in the lower frequency register. Most probably this is due to the improved F0-discrimination ability of the subjects when using the F0mod scheme. This allowed the subjects to better discriminate small note intervals in the melodies and thus more easily recognize the melodic contour. Gfeller *et al.* [2002] also found that subjects that were able to discriminate smaller F0-differences could recognize more melodies. Although F0-discrimination performance at 370.0 Hz was not different for ACE and F0mod, there is still a difference in performance of melody recognition for melodies around 370.0 Hz. Probably the subjects were able to use temporal pitch cues for the intervals below 370.0 Hz, because notes as low as 220 Hz were present in these melodies.

The subjects CI1, CI3, and CI5 perform relatively well compared to the melody-recognition performance of the subjects in the study of Kong *et al.* [2004]. In this latter study only one out of six CI subjects was able to recognize isochronous (i.e. without rhythm) melodies above chance level. However, melody recognition performance varies greatly across CI subjects [Pijl and Schwarz, 1995b; Gfeller et al., 2002; Leal et al., 2003]. Moreover, the pulse rates per channel may not have been high enough to correctly transmit temporal pitch cues in the CI-systems of the five subjects in the study of Kong *et al.* that performed close to chance level at the melody recognition task. It has been shown that the pulse rate needs to be approximately 4 to 5 times higher than the maximum modulation frequency that is to be transmitted [McKay et al., 1994]. However, in the study of Kong *et al.* [2004] the pulse rate per channel was below 300 pulses per second for three subjects (using the SPEAK strategy) and below 900 pulses per second for one subject (using the Clarion I CIS) while the frequencies of the notes of the melodies ranged from 207 Hz to 523 Hz. In a later study with other patients implanted with more recent CI devices, Kong *et al.* also found better melody recognition comparable to our results [Kong et al., 2005].

It should also be noted that the performance of subjects CI1, CI3, and CI5 corresponds to the performance of three Nucleus CI22 subjects that were asked to recognize isochronous melodies presented by solely varying the pulse rate [Pijl and Schwarz, 1995b]. The performance of these subjects is also comparable to the performance of normal-hearing subjects in recognizing isochronous melodies when only temporal cues are presented [Burns and Viemeister, 1976; Moore and Rosen, 1979].

The different notes in the melodies were not loudness balanced even though the MIDI-synthesizer generated them with approximately equal RMS. This may have lead to a loudness cue in the melody recognition. In theory, the subjects might have used this loudness cue for melody recognition. However, the effectiveness of loudness cues has been shown to be very limited for melody recognition, at least in normal-hearing subjects [Moore and Rosen, 1979].

The results of the present experiment show no difference between the melodies presented in the low register and the high register. This is in contrast with the results of Pijl and Schwarz [1995b], who found a monotonic decay of melody recognition

performance with increasing frequency register of the melodies. The melodies in that study were, however, solely presented by means of temporal pitch cues while the note differences in the present experiment could be detected using both temporal and place pitch cues. The fact that subjects CI1, CI3, and CI5 were able to recognize melodies in the high frequency register means that they were able to use place pitch cues for the task. This is definitely so for the ACE scheme where temporal pitch cues were not present because temporal modulations above 250 Hz were filtered out in the ACE scheme. Also the effectiveness of temporal pitch cues is very limited above 300 Hz [Shannon, 1983a; Zeng, 2002].

7.6 Experiment 4: Musical interval identification

In the previous experiments we showed that CI subjects could discriminate smaller F0-differences with the F0mod scheme and also that melodies could be recognized better. However, this does not mean that the pitches are correctly transmitted to the CI subjects. In this experiment we test whether the distance between musical notes (musical interval) is correctly presented to a musically trained CI subject with the ACE and the F0mod strategy.

7.6.1 Methods

Subject CI6, who was a professional organ and piano player before becoming deaf, participated in the present experiment. Interval identification was measured for the seven major intervals and the minor second (minor second, major second, third, fourth, fifth, sixth, seventh, and octave) both rising and falling in frequency. This corresponds to frequency differences of -12, -10, -9, -7, -5, -4, -2, -1, 1, 2, 4, 5, 7, 9, 10, or 12 semitones between the notes in the test intervals. The subject was presented with a reference note (either C3, F3#, or F4#, corresponding to 130.8 Hz, 185.0 Hz, and 370.0 Hz) followed by a second note separated in frequency by one of the test intervals. The subject was then asked to indicate the interval separating both notes from the list of possible intervals. The intervals were presented in blocks and each block contained the 16 test intervals for the same reference note and processed with the same scheme. Each interval was repeated twice and the order of the intervals within one block was randomized. At the beginning of each block five additional trials were presented (randomly selected from the trials in this block) to familiarize the subject with the presented processing condition prior to data collection. The results of these five trials were discarded for the analysis. Consequently, 6 different blocks (for both ACE and F0mod schemes at the three reference frequencies) consisting of 37 trials were presented to the subject. The blocks with reference frequencies 130.8 Hz, and 370.0 Hz were presented four times to the subject. The blocks with reference frequency 185.0 Hz were presented five times to the subject. Blocks were always presented alternating between blocks processed with the ACE and F0mod scheme and all blocks with the same reference frequency were presented in sequence. No feedback was presented to the subject in order to force the subject to use his intuitive sense of musical intervals.

The intervals consisted of two dotted eighth notes rendered at 120 bpm with the software synthesizer as in experiment 2 of this chapter and for the clarinet

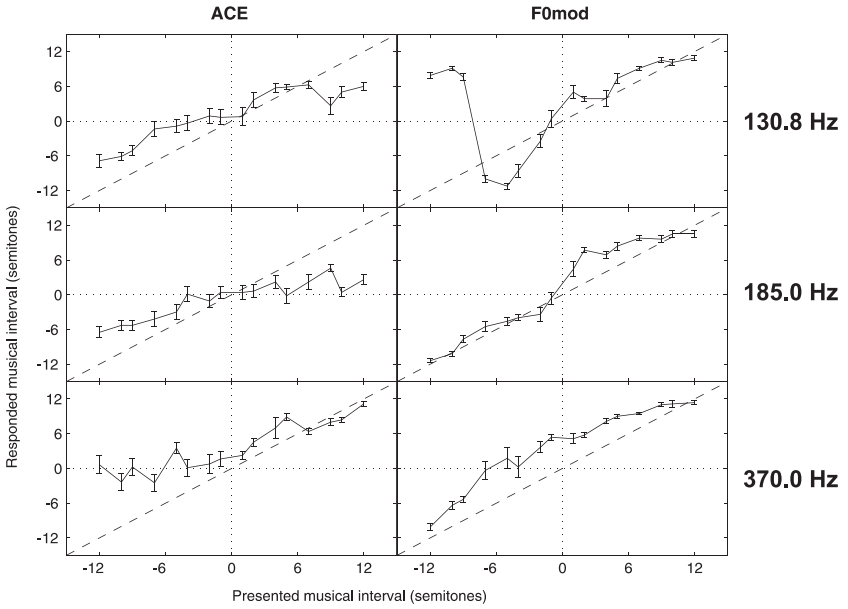


Figure 7.7: Estimates of musical pitch intervals by subject CI6 who was a professional musician before becoming deaf. The sounds processed with the ACE strategy and with the F0mod strategy are shown in the left and right hand side columns, respectively. The reference note of the two-note interval for the estimates was 130.8 Hz, 185.0 Hz, and 370.0 Hz for the top, middle, and bottom row of panels, respectively. The dashed diagonal line in every panel indicates the relation expected for musically trained normal-hearing subjects. The error bars indicate ± 1 standard error of the mean.

instrument. All notes were equalized in RMS and truncated to 500 ms. The intervals, consisting of a reference note and a comparison note with a 500 ms silent gap in between, were processed with the ACE and F0mod schemes. For the highest reference note, the upper limit of the F0-estimator of the F0-scheme was set to 750 Hz to include the highest F0-frequency (740 Hz) of the comparison note.

7.6.2 Results

The average interval estimate is shown in Figure 7.7 for each presented musical interval. The panels on the left hand side show the results for the ACE processing scheme and the panels on the right hand side show the results for the F0mod scheme. In general rising intervals were estimated as rising intervals, and falling intervals were generally estimated as falling intervals. Similarly the estimated interval became larger as the presented interval was larger. However, the subject indicated a positive difference of approximately one octave in the lowest register for the F0mod scheme for the intervals of -12, -10 and -9 semitones. This large error between the presented interval and the estimated interval is due to the incorrect estimate of F0 in the F0mod scheme in this case. The F0 of the comparison note in

these intervals were 65.4, 73.4, and 77.8 Hz which were close to or below the minimum F0 allowed in the F0 estimator. In these conditions the F0 estimator erroneously estimated the F0 at approximately 600 Hz, the maximum F0 allowed by F0 estimator of the F0mod scheme. The results corresponding to these conditions were considered outliers and have been removed from the further analysis.

The absolute errors between the estimate of the musical interval made by the subject and the presented interval were analyzed using Friedman's non-parametric two-way ANOVA. Both the processing scheme and the register had a significant effect on the size of errors made by the subject ($p < 0.001$ for both processing scheme and register). This means that the absolute error was smaller with the F0mod scheme than with the ACE scheme. Secondly, the error is larger for the higher registers than for lower registers.

The difference between the two sound processing schemes is most apparent for the larger intervals presented. For the ACE scheme large intervals (above a fifth or 7 semitones) are seldom correctly perceived. In fact, the largest perceived differences generally saturate near 6 semitones. This leads to a perceived total dynamic range of 12 semitones or less while a range of 24 semitones was presented. In contrast, for the F0mod scheme most intervals, including the larger intervals above a fifth, are estimated correctly, and consequently the perceived total dynamic range approximates the 24 presented semitones.

The notes between 185.0 Hz and 370.0 Hz for the ACE scheme (rising intervals from 185 Hz and falling intervals from 370.0 Hz) were perceptually approximately the same because the average estimate of the corresponding intervals is close to unison for each interval. In this frequency range the temporal pitch cues were probably very limited due to the processing and the place pitch cues were unable to create musical pitch differences.

The larger just noticeable differences in F0 with the ACE scheme are also reflected in a larger variance of the estimates with the ACE scheme. The average standard deviations of the estimates were 3.1 semitones and 2.1 semitones for the ACE and F0mod scheme respectively.

7.6.3 Discussion

The results of the present experiment agree with the results of the previous experiments of this chapter. In experiments 1 and 2 a smaller difference limen was found for the F0mod scheme. This created a finer perceptual grid of the musical pitches and consequently, the subject was able to more accurately denote the musical intervals. The results of this experiment also support the finding of experiment 3, in that the improved melody recognition with the F0mod scheme was most probably due to a more correct perception of the musical intervals between the notes.

The results of this experiment are related to the results of Pijl and Schwarz [1995a; 1995b] and Pijl [1997]. They measured interval labeling using the intervals of familiar melodies in three CI subjects and found that interval identification was possible by solely varying the pulse rate of the stimulus. They also found that the estimated intervals matched the intervals normal-hearing subjects would perceive

with the same frequency ratios. This result was extended to amplitude modulated pulse trains with a former piano tuner who received an implant after having become deaf [McDermott and McKay, 1997]. These results support our findings of good correspondence between presented and estimated musical interval with the F0mod scheme.

The standard deviation on the estimates was approximately 1 semitone for the three subjects of the studies of Pijl and Schwarz. This is smaller than the standard deviation for interval estimation found in the present study. The larger standard deviation found in the present study may be due to the differences in methodology and subject groups between the two studies. However, it may also be possible that variations in pulse rate may lead to clearer musical intervals than amplitude modulated pulse trains or that the place pitch cues present in the stimuli in the present experiment made the correct interval estimation more difficult. In contrast, Pijl [1997] also reported interval estimation for musical notes presented in sound-field through the subjects' speech processor and found that the subjects were unable to label the intervals correctly. The speech processing strategy used in the subjects' speech processors, SPEAK, is similar to the ACE strategy in the present study. Although the interval identification with the ACE scheme is relatively poor in the present experiment, it is better than the interval labeling performance in the report of Pijl [1997]. This difference is probably due to the low pulse rate per channel in the SPEAK strategy (below 300 Hz) which is too low to correctly transmit the modulation frequency of the beatings in the channels [McKay et al., 1994]. In the present ACE scheme the pulse rate per channel was 1800 pulses per second, allowing for modulations to be correctly perceived up to about 450 Hz.

It is also noteworthy that temporal pitch cues were only effective up to notes of approximately 185 Hz for the ACE scheme (because of the processing) and up to notes of approximately 370 Hz for the F0mod scheme (because the efficacy of temporal pitch cues decreases above 370 Hz, see experiment 2 of this chapter). Above these frequencies only the place pitch cues in the stimuli were effective. For the ACE scheme the place pitch cues were ineffective in eliciting the correct musical pitch difference between 185 Hz and 370 Hz as the two notes in the intervals were estimated to be similar in this frequency range. However, above 370 Hz the place pitch cues were able to elicit musical pitch differences that moreover also resembled the presented musical pitch differences based upon F0. A similar result is observed for the F0mod scheme above 370 Hz where place pitch cues elicit a monotonically rising musical interval percept. However, in this case the musical interval was consistently overestimated. McDermott and McKay [1997] have also reported that interval estimation is possible using solely place pitch cues and that the perceived interval rises monotonically with increasing spatial separation between the stimulation sites.

7.7 General discussion

7.7.1 Implications for CI sound processors

The present study indicates that providing explicit modulation of the channel envelopes at F0 with large modulation depth and in phase across channels, as in the F0mod scheme, improves (musical) pitch perception in CI subjects. This result is obtained from a comparison to conventional sound processing schemes where temporal pitch cues related to F0 are only present through beatings (of relatively shallow depth) originating from the interactions of the harmonics falling within the same analysis filter. The benefit found in the present study was also larger than the benefit found by Green *et al.* [2004] with a sound processing scheme similar to the proposed F0mod scheme. This was most probably due to differences in the nature of the stimuli.

Apart from the improved F0-discrimination with the F0mod scheme, the modulation depth at F0 is unaffected by reverberation or background noise (as long as the F0-estimation is performed correctly). This means that the temporal pitch cues remain maximally salient even under adverse real-life circumstances. With the ACE scheme, or equivalent schemes, the modulation depth of the F0 related beatings is greatly reduced by reverberation or background noise, and consequently the efficacy of the temporal pitch cues declines. These benefits in pitch perception justify the implementation of the F0mod scheme or a similar scheme in a portable sound processor, to evaluate its performance under every-day-life circumstances. The proposed scheme is also compatible with other improvements to CI sound processing schemes that were recently developed mainly for improved speech and loudness perception [Geurts and Wouters, 1999; Vandali, 2001; James et al., 2002; McDermott et al., 2003], because these schemes do not interfere with relatively fast channel envelope variations above 100 Hz.

However, a number of issues still need to be resolved before an explicit F0 modulating scheme can be used clinically in the CI population. Firstly and most importantly, it must be assessed whether the proposed sound processing scheme does not have an adverse effect on speech intelligibility. Previous studies have shown that speech perception is improved in most subjects when the stimulation rate per channel is raised above 150 Hz [Fu and Shannon, 2000] or 250 Hz [Skinner et al., 2002a; Skinner et al., 2002b]. In the proposed scheme, the effective stimulation rate may be reduced to frequencies in the range of 100 to 300 Hz and consequently a negative effect on speech perception may be expected. A preliminary study of vowel identification with the scheme proposed by Green *et al.* [2004] showed that vowel recognition by CI subjects with the explicit F0-modulating scheme was worse than with standard CIS [Macherey, 2003]. Further evaluation of speech perception with the proposed scheme is necessary. In case a negative effect is found on speech perception with explicit F0-modulating sound processing schemes, such schemes may still be useful in a number of circumstances such as listening to music or perception of tones in tonal languages. In such circumstances, the subjects sound processor may be switched (automatically or manually) from the standard sound

processing scheme, optimized for speech perception, to the explicit F0-modulating scheme, or to a combined scheme in between these extremes.

A second issue that needs to be resolved before clinical application of the F0-modulating schemes has a more technical origin. A robust F0-estimation is essential for the correct functioning of the proposed scheme. It has to be computationally inexpensive and has to have a low latency. Combining all these features in adverse circumstances has been proven difficult in the past.

It should be noted that although the enhanced envelope pitch cues in the F0mod scheme resulted in improved music perception, these cues are weak compared to pitch cues in normal-hearing subjects. Optimally, the salience of pitch cues related to resolved harmonics should be transmitted to CI subjects.

7.7.2 Mechanisms of pitch perception

It has been shown that purely temporal pitch cues in NH listeners and temporal pitch cues in CI listeners can convey musical pitch [Burns and Viemeister, 1976; Moore and Rosen, 1979; Pijl and Schwarz, 1995b]. However, there is an ongoing debate whether spectral pitch or place pitch meets the strict definition of musical pitch. Some researchers have argued that this is not the case [Moore and Carlyon, 2005]. Our results indicate that also place pitch cues can convey musical pitch. Firstly, three of the four subjects were able to recognize melodies without rhythm cues in the high register with the ACE strategy. In this condition, the lowest note was around 220 Hz so for all notes temporal pitch cues were absent because of the envelope filtering in the ACE strategy. This means that most probably the subjects recognized the melodies purely based on place pitch cues. Secondly, in the interval estimation experiment the musically trained subject CI6 was able to approximately label the musical intervals for notes above 370 Hz. In this frequency region, temporal pitch cues are assumed to be completely absent in case of the ACE scheme, and at least to be less effective in case of the F0mod scheme.

Our results appear to indicate that place pitch cues can be used for musical pitch in some particular tasks, similar to temporal pitch cues. This does not mean that both cues elicit the same musical pitch percept. Most likely this is even not the case, because of their psychophysical [Tong et al., 1983; McKay et al., 2000] and physiological [Warren et al., 2003] independence.

7.8 Conclusion

In this chapter a new sound processing scheme for cochlear implants (F0mod) was proposed to enhance pitch and music perception by CI subjects. The scheme was designed to optimally present both the temporal and place pitch cues to the CI subjects. Temporal pitch cues were presented as clear and as unambiguously by modulating the envelopes of all channels sinusoidally at the pitch frequency with 100% modulation depth and in phase across all channels. The place pitch cues were presented linearly compressed along the cochlea.

With the new F0mod scheme music perception was found to improve significantly with respect to the current most often used sound processing strategy for Cochlear Nucleus recipients. This was evidenced by three benefits:

1. F0-discrimination was three times better with the F0mod strategy compared to the standard strategy for musical notes of five different instruments when the F0 was below approximately 250 Hz.
2. Subjects were able to recognize more melodies (without any rhythmic cues) with the F0mod sound processing scheme.
3. The perceived pitch differences between musical notes were more in accordance with the musical scale as perceived by normal-hearing subjects. Consequently, musical pitch differences are more accurately presented with the new sound processing scheme, and thus provide the subjects with a more correct contour of the melody.

There was, however, no clear benefit with the new sound processing scheme for F0-discrimination at 370 Hz and for interval labeling above 370 Hz where only place pitch cues were available. This means that there was no significant difference in perception of the place pitch cues between the new scheme and the standard scheme, at least for these conditions.

These findings indicate that the presented F0mod scheme is very likely to improve CI subjects' music perception or speech understanding in tonal languages.

Future research will implement the F0mod scheme in a wearable processor for subjects for take home experience. This implementation should include a robust F0 estimator for noisy conditions and an extension of the F0mod scheme for non-harmonic sounds. Tests will include speech perception testing. Speech intelligibility may deteriorate with the F0mod scheme because of the lower effective stimulation rate, or speech intelligibility may improve because of the improved voice perception.

7.9 Appendix: The relation between spectral and place pitch

The difference in spectral pitch $\Delta spectral$ between two acoustical stimuli in a normal-hearing subject is given by the difference between the centroids of the acoustically evoked excitation patterns of the two stimuli [Zwicker and Fastl, 1999]:

$$\Delta spectral = C_2^{acoust} - C_1^{acoust} = \frac{\int x M_2^{acoust}(x)\,dx}{\int M_2^{acoust}(x)\,dx} - \frac{\int x M_1^{acoust}(x)\,dx}{\int M_1^{acoust}(x)\,dx} \tag{7.6}$$

where C_1^{acoust} and C_2^{acoust} are the centroids of the acoustically evoked excitation patterns of stimuli 1 and 2 respectively, x is the distance along the basilar membrane, and $M_n^{acoust}(x)$ is the acoustically evoked excitation pattern of the n-th stimulus.

The difference in place pitch $\Delta place$ between two electrical stimuli in a CI subject is given by the difference between the centroids of the electrically evoked excitation patterns of the two stimuli (see chapter 5):

$$\Delta place = C_2^{elec} - C_1^{elec} = \frac{\int x M_2^{elec}(x)\,dx}{\int M_2^{elec}(x)\,dx} - \frac{\int x M_1^{elec}(x)\,dx}{\int M_1^{elec}(x)\,dx} \tag{7.7}$$

where C_1^{elec} and C_2^{elec} are the centroids of the electrically evoked excitation patterns of stimuli 1 and 2 respectively, x is the distance along the basilar membrane and $M_n^{elec}(x)$ is the electrically evoked excitation pattern of the n-th stimulus.

When the sound processing scheme linearly warps the resulting excitation pattern along the basilar membrane:

$$M_n^{elec}(x) = M_n^{acoust}(a\,x + b), \tag{7.8}$$

then it can be shown that:

$$C_n^{elec} = \frac{1}{a}\left(C_n^{acoust} - b\right), \tag{7.9}$$

and consequently

$$\Delta place = \frac{1}{a}\Delta spectral\ . \tag{7.10}$$

Chapter 8

Conclusions and future research

The first three sections of this chapter summarize the conclusions of this thesis and section 8.4 provides some suggestions for future research.

8.1 The mechanisms for pitch perception in CI subjects

Compared to normal-hearing persons, CI subjects very poorly perceive pitch and pitch differences. Normal-hearing persons can distinguish two sounds with a very small pitch difference of 0.1 to 0.2 semitones. In contrast, CI subjects (typically) only notice pitch differences of about 7 semitones with the current sound processors (see Figure 8.1). The performance of CI subjects is thus approximately 30 times worse than the performance of normal-hearing persons, and it is too poor to allow for, for instance, music enjoyment. It was the goal of this thesis to reveal the mechanisms of pitch perception in cochlear implants and use this acquired knowledge to optimize the sound processing in CI systems for better pitch perception.

In order to understand, and possibly improve, pitch perception by patients with a CI, a number of psychophysical experiments were carried out. These investigations assessed how the two pitch mechanisms in CI subjects (place pitch and temporal pitch) contribute to the perceived tone height and to F0-discrimination.

First we developed in chapter 2 a research platform that allows to set-up and run arbitrary experiments with normal-hearing persons and with CI subjects. The

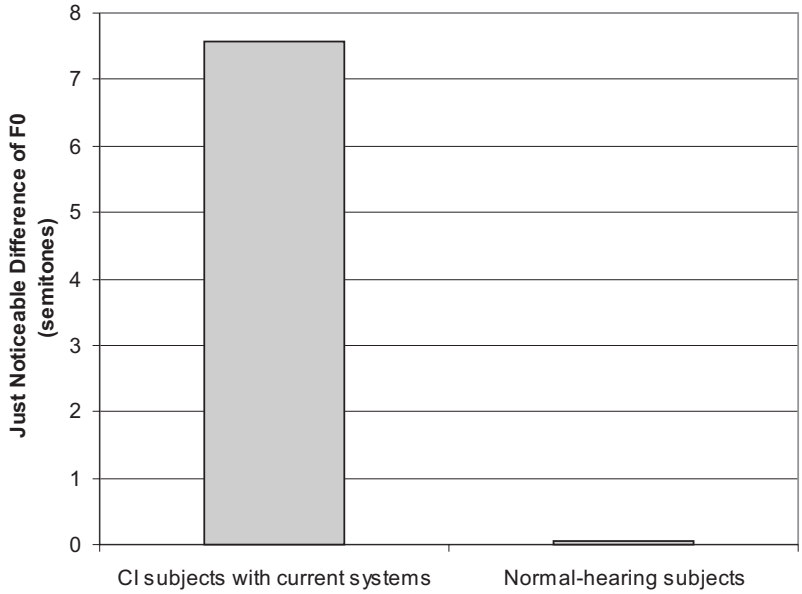


Figure 8.1: Pitch discrimination performance for CI subjects with current systems, and for normal-hearing persons. Smaller values indicate better performance. The results of the CI subjects are the mean pitch discrimination performance from 46 CI-subjects with various devices taken for piano notes from Gfeller et al. [2002]. The results of the normal-hearing persons are the average of the subjects and of the different vowels from [Flanagan and Saslow, 1958].

graphical user interface presented to the subjects can also be animated to allow testing of young children from the age of 2.5 years. This research platform was used in all experiments performed as part of this thesis as well as in a number of other studies.

8.1.1 Place pitch

Place pitch is known to be related to the place of stimulation along the cochlea. In chapter 5, our results indicated that more specifically place pitch cues are related to the centroid or center of gravity of the electrical stimulation pattern along the cochlea for complex multi-channel stimuli. This finding is consistent with the finding of chapter 3 that place pitch does not vary with spatial overlap, but varies with the degree of shift in the stimulation pattern. The proposed centroid model for place pitch is linear, and independent of overall signal intensity.

Based upon the centroid model, subject specific pitch perception results can be predicted by adjusting a single parameter, the subjects' place pitch sensitivity. This parameter can easily be measured through, for instance, an electrode ranking

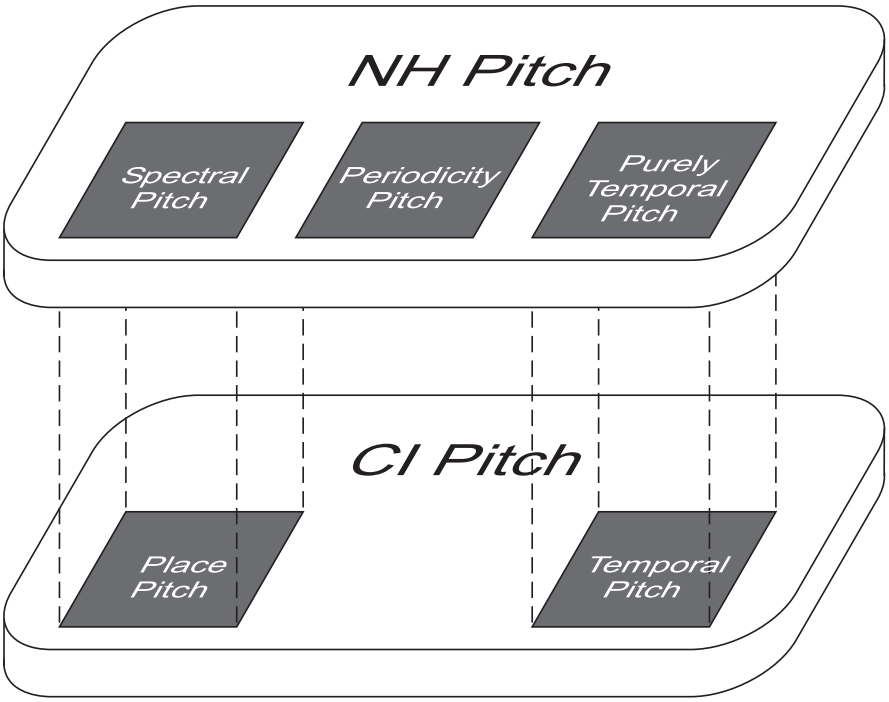


Figure 8.2: Schematic overview of the pitch mechanisms in normal-hearing persons and in CI-subjects. The pitch mechanisms for normal-hearing persons are shown in the upper part and the pitch mechanisms for the CI-subjects are shown in the lower part. The correspondences between mechanisms of normal-hearing and CI subjects are indicated by the dashed lines.

experiment. Predictions also hold for pitch perception with complex multi-channel stimuli.

The centroid model also implies that place pitch is determined by the overall spectral shape of the processed sound. Place pitch is therefore more closely related to the timbre of a sound (like formants or resonance frequencies) than to the pitch or F0 of the sound. More specifically, place pitch in CI subjects appears to be related to the sharpness of a sound, sometimes also denoted as brightness or spectral pitch for normal-hearing persons (see the dashed lines in Figure 8.2).

This last finding somewhat questions the validity of the definition of place pitch as a pitch sensation because place pitch does not relate directly to the F0 of a sound which is usually used to convey melody in music. Stricter definitions of pitch require pitch sensations to possibly convey melody [Plack and Oxenham, 2005a]. However, in chapter 7 we show that place pitch cues can evoke a sense of musical pitch and that place pitch cues can be used for melody recognition similar to F0-differences and they can also be used for musical interval labeling.

In chapter 6, using an acoustic model of CI sounds in normal-hearing subjects, we measured place pitch sensitivity for varying slopes of the stimulus pattern. The

results indicated that place pitch differences were more easily detected for narrow stimulus patterns with steep spectral slopes. These results can most probably be extrapolated to CI-subjects, and infer that place pitch sensitivity is related to the spread of excitation along the cochlea, where more focused stimulation leads to better place pitch sensitivity. Note, that the place pitch sensitivity for the different subjects in the experiments of this thesis was relatively good (below 0.75 mm) even though in all conditions for monopolar stimulation was used. Together with the previous statement, this implies that monopolar stimulation is capable of exciting neural excitation patterns with relatively steep slopes.

8.1.2 Temporal pitch

Temporal pitch in CI subjects is related to the modulation frequency of the temporal envelope of the electrical stimulation pattern. In chapter 5, we measured the effective temporal pitch cues for different stimuli. The analysis of the results indicated that the temporal pitch mechanism provides the clearest cues to the subject when the envelopes are modulated at the same frequency and in phase across the different channels. We also showed that the temporal pitch cues are more salient when the modulation depth of this 'in-phase' modulation is deeper.

The fact that channels which are modulated asynchronously elicit less clear cues for pitch discrimination is most probably related to channel-interaction. Because of the current spread along the cochlea with electrical stimulation, broad excitation patterns are activated. As a consequence, the depth of the dips in time in the neural excitation elicited by a single modulated channel can be reduced because a neighboring channel might excite the same population at the same time instant.

Our results also indicate that the temporal pitch cues are most salient when only a single modulation frequency is present over the different channels. At least two reasons exist for this. Firstly, when multiple modulation frequencies exist on neighboring channels, the resulting modulation depth of the neural excitation may be reduced because of the channel interaction. Secondly, the complex pitch perception mechanism, or periodicity pitch mechanism, is impaired when the normal tonotopic pattern is disturbed [Oxenham et al., 2004]. This is almost always the case in CI systems.

We showed that the poor performance of the CI subjects with current sound processing schemes does not originate from the fact that the current filter banks do not resolve the individual harmonics as in the normal ear. The fact that the BUT condition in chapter 5 provided the least temporal pitch cues highlights this point because the BUT filter bank resolves the lower harmonics.

8.1.3 The combination of place and temporal pitch cues

We showed that place pitch cues and temporal pitch cues are integrated into a single 'global' pitch percept when the subjects are forced to do this in a pitch ranking task (see chapter 4). For consistent pitch cues, this leads to a more salient combined 'global' pitch sensation. For inconsistent pitch cues, where one cue indicates a falling pitch difference and one cue indicates a rising pitch difference, both mechanisms are able to cancel each other. This does not mean however, that place

pitch and temporal pitch sensations are interpreted along the same perceptual dimension. Place and temporal pitch mechanisms lead to independent pitch percepts which can be integrated (according to an 'adjusted' Euclidean distance measure) when the cues are ambiguous by themselves.

We also showed that because of this combination both place and temporal pitch cues contribute to F0-discrimination by CI subjects. In general, both cues contribute to the pitch CI subjects hear when listening through their sound processors.

We also found a remarkable difference in preference of pitch cue between subjects. Some subjects relied more on place pitch cues, other subjects relied more on temporal pitch cues. No clear reason for this difference was found.

8.2 F0-discrimination by CI subjects

8.2.1 Contributions of place pitch and temporal pitch to F0-discrimination

We showed that both place pitch and temporal pitch cues contribute to the general pitch CI subjects perceive and that they use both cues to discriminate F0-differences (see chapters 5 and 7).

From the results of chapters 5 and 7, F0-discrimination solely based upon place pitch cues is rather weak (with JNDs around 6 to 12 semitones). F0-discrimination only based upon place pitch cues also varies strongly across measuring conditions even within the same subjects. This variance is caused by the non-monotonic behavior of the place pitch cues coming from a close interaction between spectral characteristics of the sound (such as formants and F0) and the analysis filter bank in the CI. This interaction is a direct consequence of the centroid model for the place pitch mechanism. The interaction in perception between F0 and spectral shape does not exist in normal-hearing persons, at least not to the same extent.

The poor performance and the interaction with the spectral shape implicate that solely place pitch cues might be hard to use for F0-discrimination in daily life for CI subjects. However, we also showed that in a number of cases place pitch cues are the only cues effectively presented to the subjects (for instance for F0s above 300 Hz). It is also worth noting, that in chapter 4 a difference was found between subjects. Some subjects relied more on place pitch cues and some relied more on temporal pitch cues. This highlights the importance of the place pitch cues for sounds with F0s above 300 Hz and for subjects that tend to rely more on place pitch cues than on temporal pitch cues, despite the limited performance when pitch perception is only based upon place pitch cues.

The addition of temporal pitch cues greatly improves F0-discrimination performance by CI subjects. In other words, F0-discrimination is best when also temporal pitch cues are effectively presented to the CI subjects. With the inclusion of optimally presented temporal pitch cues F0-discrimination was possible up to 1 or 2 semitones (either for anechoic harmonic stimuli with the current sound processors or for all harmonic stimuli with the new F0mod strategy). This means that temporal pitch cues

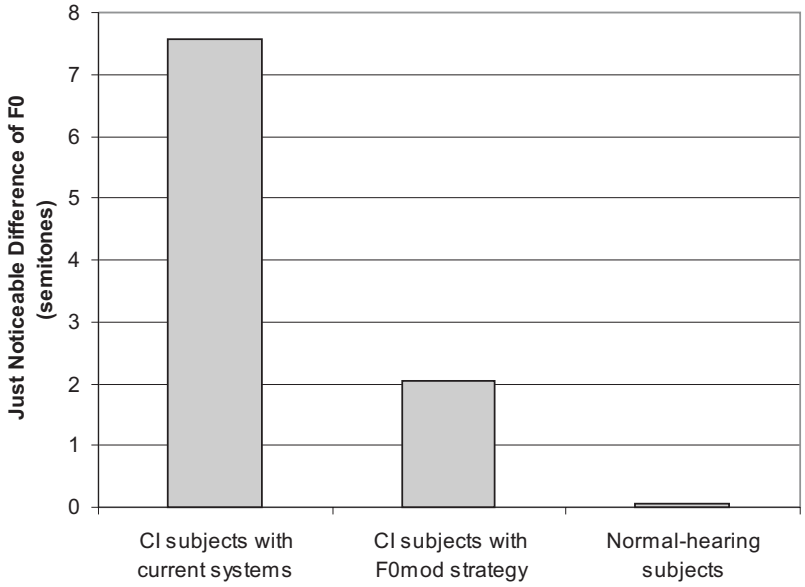


Figure 8.3: Pitch discrimination performance for CI subjects (with current systems and with the new F0mod strategy) and for normal-hearing persons. The figure is based upon Figure 8.1 with inclusion of the results for the CI subjects with the new F0mod strategy. The results of the CI-subjects with the F0mod strategy are derived from the results of chapter 7 averaged over the 6 subjects and all conditions (vowels and musical notes).

are the major contributor to F0-discrimination in CI subjects. Temporal pitch cues can also be used in daily life because temporal pitch is perceived independently from other sound properties. Unfortunately, the limit of resolution with temporal pitch cues appears to be encountered because JNDs is located around 1 to 2 semitones and temporal pitch cues are only effective below 300 Hz.

8.2.2 Influence of sound processing on F0-discrimination

In chapter 5, we measured the performance of CI subjects for F0-discrimination (or pitch discrimination) for the current sound processing strategies. By varying the filter bank and the envelope extraction mechanism we were able to enhance or deteriorate the usable pitch cues for the CI subject. The envelope extraction mechanism (whether based upon half-wave rectification or based upon the Hilbert transform) did not significantly affect F0-discrimination as long as the highest modulation frequency passed by the envelope detector was high enough to allow for temporal pitch cues (up to 300 Hz).

The comparison of the results for the four filter banks tested showed that filter banks with higher resolution in the F0-frequency register transmitted the place pitch cues.

In contrast, for filter banks with lower resolution the temporal pitch cues were transmitted more clearly. As a consequence of this inconsistency, it was impossible to optimize the transmission of both place and temporal pitch cues by varying the specifications of current sound processing schemes.

In chapter 7, we therefore developed a new sound processing scheme (F0mod) that is designed to optimally transmit both the temporal and place pitch cues to the CI subject. In this new sound processing strategy, temporal pitch cues were presented using explicit modulation of the channel envelopes at F0 because in the previous chapters we showed that temporal pitch cues are maximally effective for synchronous and deep modulation. Place pitch cues were presented linearly warped along the cochlea. The results of the experiments showed that with the F0mod strategy the CI subjects could discriminate pitch differences of only 1 to 2 semitones. This is about three times better than with the current sound processing strategies (see Figure 8.3). Unfortunately, this improvement appears to be restricted to fundamental frequencies up to approximately 250 Hz. Above this frequency, the temporal pitch cues are less effective, and the F0-discrimination with the new scheme is comparable to the performance with current sound processing schemes.

The improved F0-coding in the F0mod sound processor does not only lead to improved F0-discrimination for vowel-like stimuli and for musical notes. We also showed that with the F0mod strategy more melodies are recognized by the CI subjects and that there is a closer correspondence between presented and perceived musical pitch differences. In summary, pitch and melody perception by CI subjects are improved with the new F0mod sound processor.

8.3 An acoustic model for pitch perception by CI subjects

In chapter 6, we developed a simulation of the sound CI subjects hear. We presented this CI-simulation to normal-hearing listeners in order to infer knowledge on the pitch sensation experienced by CI users. CI-simulations can have a didactical or clinical use: for instance, using CI-simulations, the parents of a deaf child receiving a CI can experience the relatively poor quality of the current CI systems. However, the most important benefits of CI-simulations (or acoustic models) are that they provide us a tool to investigate parameters that can not be varied in CI users (e.g. current spread) and without the complicating factors inherent to CI research coming from the various etiologies, durations of deafness, length of implant experience, etc.

In chapters 4 and 6, an acoustic model was used with normal-hearing subjects to investigate pitch phenomena in CI subjects. The model is based upon a noise-band vocoder. Previous studies had shown the accuracy of noise-band vocoders as acoustic models for speech understanding in CI subjects. In the model used in this thesis, place pitch cues were altered by changing the spectral shape of a noise band (center frequency and spectral slopes). Temporal pitch cues were inserted through amplitude modulation of this noise band. These design specifications were based upon the correspondence between place pitch in CI and spectral pitch in normal-hearing persons, and upon the correspondence between temporal pitch cues in CI

and envelope (purely temporal) pitch cues in normal-hearing persons (see Figure 8.2).

In chapter 6, we compared results obtained in CI subjects with results obtained using this acoustic model in normal-hearing persons. We showed that it is possible in the acoustic model to improve the subjects' sensitivity to place pitch changes by increasing the steepness of the spectral slopes of the noise bands. When the spectral slope of the noise band was adjusted so that the normal-hearing subjects' sensitivity to place pitch differences matched that of CI subjects, the model accurately predicts place pitch cues even for complex multi-channel stimuli.

We also showed that, in line with previous studies, temporal pitch cues are less effective using the model when the noise-band carrier has a low center frequency (below 1 kHz). For noise bands with low center frequencies, the peripheral auditory filters limit the effective perceived modulation depth, and consequently the temporal pitch sensitivity. Care has thus to be taken when choosing the carrier frequencies for a vocoder. A noise-band vocoder with high center frequencies for the resynthesis filters (e.g. mimicking a shallow insertion depth of the simulated electrode array) may overcome this effect of the peripheral auditory filters.

In summary, the noise-band vocoder can be used as an effective acoustic model to study pitch perception in CI-subjects. However, care has to be taken to adjust the spectral slopes of the noise-band carriers (to mimic place pitch sensitivity) and to set the center frequencies of the noise-band carriers above approximately 1 kHz (to lift the ringing effect of the peripheral filters in the normal ear).

Two important remarks have to be made concerning acoustic models or CI-simulations. Firstly, I want to highlight that acoustic hearing is fundamentally different from electric hearing and results obtained with an acoustic model always have to be validated with CI subjects. Secondly, the sound experience CI-simulations evoke in NH listeners is still very different from the sound CI users hear. For instance, the CI-simulations presented in this thesis sound very noisy, but CI users generally describe their sound experience as very metallic or computerized. The CI-simulation has to be interpreted as a sound of which the quality is similar to the (limited) quality of the sound heard by CI users, but both sounds are not equivalent.

8.4 Indications for future research

8.4.1 Research platform

The research platform that we developed in chapter 2, APEX, allowed to set-up and run all experiments described in this thesis. Apart from this thesis APEX is recently also used in a number of other studies by different research groups around the world (Pretoria, Zurich, Birmingham, Auckland, Utrecht, Amsterdam, Cambridge, etc). Although it is already very flexible and powerful APEX still possesses a number of limitations. A first limitation is that a number of testing procedures are not implemented yet within the research platform. These procedures include scaling experiments (although this is already feasible with open or closed set identification),

adaptive closed set identification, speech in noise testing with an adaptive signal to noise ratio.

Secondly, because APEX is designed to output to only one or two channels, APEX is not suited for localization experiments assessing binaural hearing. Localization experiments typically require a large number of output channels (more than 8). Currently the research platform is also incapable of driving different output devices simultaneously. This prevents testing of bimodal stimulation (CI and acoustic hearing) or of binaural CI subjects when a direct input to the device is required.

A third limitation of the present state of APEX is that it discards the response time of the subjects. For a number of experiments it might also be interesting to record the response time of the subjects because the response time is often related to the subjects' performance. At present the response times are not measured because this requires a more accurate timing of both the sound output and the subjects input which can be hard on the Windows operating system.

Another extension to APEX would be the definition and implementation of a new file format. The increase in functionality of APEX, which will most probably even increase further in the future, interferes with the ease-of-use of the file format for experimental set-up and results as it is defined at present (in text files). In order to retain the user friendliness of APEX and still allow for more functionality it may be required to define a new experiment definition format based upon XML which than could be edited by a custom developed graphical user interface.

In order to be able to access relatively fast a very large population of possible subjects, it may be interesting to perform psychophysical test through the internet. An example of an auditory psychophysical test for speech understanding in noise is already implemented by the Audiological Centre of the free university (VU) of Amsterdam (http://www.hoortest.nl).

8.4.2 Pitch mechanisms in CI

In chapter 3, place pitch sensitivity was measured for multi-channel stimuli. However, the place pitch was only compared for stimuli containing the same number of active channels and having a flat spectral profile. It might be interesting to compare the place pitch of stimuli that differ in the number of active channels and have different spectral profiles. This would allow validation of the centroid model. For example, if place pitch was primarily determined by the apical or basal edge of the excitation pattern, then the stimulus with more electrodes would have a different pitch than the one with fewer electrodes even if they both had the same centroid. Because the apical edge and the centroid of the stimulation patterns were correlated in the stimuli used in the experiments of this thesis, it was hard to segregate the influence of both properties on the perceived place pitch.

The centroid model was also inferred from experiments with stimuli delivered in monopolar mode. Because monopolar mode excites broad neural responses, the resulting multi-channel neural excitation profiles were relatively smooth. It can be questioned whether the centroid model would still hold for stimuli consisting of a number of very narrow spectral excitation patterns.

From the acoustical model there was also evidence that place pitch sensitivity improved with steeper edges of the stimulation pattern. It would be worth verifying this feature in CI subjects. Sharper excitation profiles in CI subjects through new modes of stimulation (current steering or alternative electrical waveforms) would than enhance place pitch sensitivity in CI subjects. Such an improvement would result in an increase of the place pitch cues for F0-discrimination.

In chapter 5, we found that temporal pitch cues are more salient when the modulation depth of the envelope is increased. It is however unknown how other features of the modulating waveform affect temporal pitch saliency. For example, the 'peakedness' of the waveform may have an effect of the resolution, and similar the duty cycle of the modulating waveform. The waveform used in the F0mod strategy could be optimized to provide the best rate discrimination.

We showed that place and temporal pitch cues can be integrated even when they are conflicting (see chapter 4). However, most probably temporal and place pitch cues are not always integrated because they remain separate and independent sensations. At present it is unknown where the boundary lies between the integration of (ambiguous) cues and the perception of two independent sensations.

One of the most striking results of chapter 4 was also that there exists a difference between subjects in their 'preference' for pitch mechanism (place pitch or temporal pitch). For the understanding of pitch mechanisms and auditory perception in general the reason for this difference has to be revealed. This dichotomy in subjects' behavior might reconcile the spectral (or place) views with the temporal views on auditory perception. However, the cause of this between-subject difference has to be investigated further. This can be done by, for instance, assessing the subjects sensitivity to pitch changes (both spectral and temporal), or the physiologic processing of pitch evoking stimuli using frequency following responses or brain imaging. Musical training appears not to have an effect on the 'preference' of the subjects because both the 'place-preferring' group of subjects and the 'temporal-preferring' group of subjects contain trained musicians.

8.4.3 Acoustic model of CI

The acoustic model based upon the CISIM vocoder developed in chapter 6, allowed for a relatively good match between results obtained from CI subjects and results obtained with the model from normal-hearing persons. Noise-band vocoders (with adequately designed resynthesis filters) thus appear usable as an acoustic model for pitch perception in CI subjects. We must however remember the large and fundamental difference between the stimuli of the two modes. The carriers in the vocoders are stochastic in nature while the carriers in the CI system are deterministic pulse trains. This difference may have also contributed to the difference in the effectiveness of the temporal pitch cues between the CI results and the vocoder results.

In order to improve the quality of acoustic models, future acoustic models have to reduce the difference between the neural excitation pattern elicited by the acoustic model in normal-hearing persons and the neural excitation pattern elicited by the CI in implanted subjects. As a first step, the design of the filter shape of the CISIM

filters should take into account the effect of the peripheral filtering (for example upward spread of masking). In a next step the acoustic model should also take into account the difference in transformation for mechanic movement to neural signals and the transformation from electric current to neural excitation. This would result in an acoustic model that consists of two parts: firstly, a complete model of neural excitation based upon an electric stimulation pattern; secondly, a reverse model of the neural excitation of the auditory periphery of the normal ear. An example of such a model has been recently developed by Huang et al. [2004].

In the minimization of the difference between electric excitation pattern and the acoustic excitation pattern, most currently used acoustic models only take into account the magnitude of the rough spectral and temporal envelope. It may be better to include phase information of the excitation pattern because, for example, the phase relations between modulations of adjacent neural regions are important.

With the advent of an increasing population of CI subjects with residual hearing and also CI subjects with CIs in both ears, it may also be interesting to extend the present acoustic model to allow for testing bimodal (combination of residual hearing and a CI) and binaural hearing.

8.4.4 Evaluation of the F0mod strategy

In this thesis we developed a new sound processing strategy, F0mod (see chapter 7), designed to enhance pitch perception in CI subjects. Although the in-lab results were extremely promising (pitch discrimination was three times better than with current sound processing strategies), a number of aspects still have to be investigated. For instance, the effect of the proposed sound processing scheme on speech perception has not been studied yet. A negative effect of the F0mod scheme on speech intelligibility would seriously harm a successful commercialization of the new sound processing strategy. To asses this and other aspects (such as training, listening in difficult circumstances, and sound appraisal) the F0mod scheme will be implemented in a wearable sound processor for the CI subjects to take home. In this way, users will be able to experience the new F0mod scheme for several weeks in daily-life conditions.

Two problems still have to be addressed for the implementation of the F0mod scheme in a wearable platform. Firstly, the F0mod scheme requires a robust F0 estimator that also works in difficult listening situations (high background noise, multiple interfering harmonic sources). Moreover, the F0 estimator has to have a low latency and a low computational cost. It has been proven in the past that the simultaneous application of those three requirements can be challenging. Secondly, the F0mod scheme is currently only defined for harmonic input sounds and thus has to be extended to include non-harmonic sounds. Although various options are feasible, the best solution has to be selected taking into account appropriate transitions between harmonic and non-harmonic sounds.

The improved F0-discrimination using the F0mod scheme may also enhance other aspects of hearing. For instance, it might be worth assessing tonal language perception using the new sound processing scheme because it is very probably that tone recognition will be improved. For the demographically very large, more than

1.3 billion people, and economically fast growing market in South-East Asia a CI system optimized for tonal languages can be a valuable asset for each commercial competitor in the CI sector.

And lastly, the F0mod scheme might improve speech perception when listening with multiple interfering speakers. Because of the improved voice perception, it may be easier to track the speech of the target speaker. In normal-hearing and moderately hearing-impaired subjects good F0-discrimination is known to aid in separating simultaneous speakers.

8.4.5 Periodicity pitch

It should be noted that, although pitch discrimination was three times better with the F0mod scheme (around 1 to 2 semitones), this performance is at least a factor of 10 worse than the performance of normal-hearing persons (0.1 to 0.2 semitones), see Figure 1.3. This means that pitch perception in CI subjects remains a problem, albeit less prominent with the F0mod scheme. Most probably this is mainly caused by our current inability to correctly trigger the periodicity pitch mechanism in CI-subjects. Consequently, CI-subjects must rely on purely temporal pitch cues. It is known from other studies in normal-hearing persons that purely temporal pitch cues are less salient than the periodicity cues. Using merely temporal pitch cues the limit for pitch discrimination appears to be reached and is about 1 semitone.

For future CI systems trying to improve pitch perception, the goal would thus be to facilitate the periodicity pitch mechanism in CI subjects. Unfortunately, the exact process(es) underlying the periodicity pitch mechanism are incompletely understood. The results of this thesis and other studies [Oxenham et al., 2004; Plack and Oxenham, 2005b] have shown that the periodicity pitch mechanism relies on a spatio-temporal feature (and not a purely temporal or a purely spatial feature) of the neural excitation pattern resulting from the second to fifth harmonic. This spatio-temporal feature is sometimes also termed ‘fine structure’ because it is related to neither the temporal envelope nor the spectral envelope. However, as already stated, the exact properties of this spatio-temporal feature are currently unknown.

Most probably, the temporal resolution required for this feature is already provided by current CI-systems because the electrically stimulated auditory nerve synchronizes very accurately to its stimulus. This is for instance evident from gap-detection thresholds in CI-subjects (which can be better than in normal-hearing persons). However, the required spatial pattern for the periodicity pitch mechanism is probably not feasible yet with current CI systems for two reasons. Firstly, the spatial resolution offered by the present CI systems is most probably insufficient. Because of the high conductivity of the intra-cochlear fluid the injected current spreads almost along the entire cochlea. As a consequence, interference exists between adjacent channels. In contrast, excitation by acoustic tones (a pure tone or a resolved harmonic) is very restricted in space. Secondly, where the place of excitation can vary continuously in the normal ear according to the frequency of the stimulus, the spatial excitation pattern in current CI systems varies in discrete steps because the activated electrode determines the excited neural population.

In order to be able to elicit the same spatio-temporal excitation pattern in CI-subjects as in normal-hearing persons, more spatial resolution is thus necessary. Recent investigations show that alternative electrode placements (radial bipolar) or alternative electrode configurations (tri-polar, current steering) might improve spatial selectivity [Snyder et al., 2004]. Current steering has the additional benefit of providing a continuous spatial dimension. Other promising investigations might be the use of 'alternative' wave-forms instead of symmetric biphasic pulses as they are currently used in CI systems [van Wieringen et al., 2005].

It may not be necessary to present an exact match of the acoustical excitation pattern in every detail through a CI system in order to enable the periodicity pitch mechanism. For instance, enhancing the spatial resolution could allow for preservation of phase-relationships between adjacent neural regions. At present these phase-relationships are lost since all auditory nerve fibers tend to fire 'in-volley'. Such relatively simple aspects of the neural excitation pattern may already be sufficient for the periodicity pitch mechanism.

However, the suggestions above are speculative because, as already mentioned, the periodicity pitch mechanism is incompletely understood at present. Understanding the periodicity pitch mechanism would require knowledge of all neural structures up to regions adjacent to primary auditory cortex because there pitch appears to be interpreted [Griffiths et al., 1998; Griffiths et al., 2001; Patterson et al., 2002; Krumbholz et al., 2003; Penagos et al., 2004]. At present only the physiology of the auditory system up to the peripheral auditory nerve is relatively well understood. Higher auditory centers are still a matter of investigation and results are complex because of the large number of neural inter-connections and because of the large number of different types of neurons. Both complicating factors are already found in the cochlear nucleus (the first stage of the central auditory system). Further research in the field of neurophysiology, neuro-imaging and modeling may unravel the mechanisms behind the periodicity pitch mechanism.

There might also be another viewpoint to facilitate the periodicity pitch mechanism in CI subjects instead of trying to replicate the fine structure of the neural excitation pattern found in normal-hearing subjects. Some theories exist that the periodicity pitch might be learned through exposure to sounds [Shamma and Klein, 2000]. If this is the case, than children that are born deaf and are implanted at a young age (during their 'sensitive period') may acquire periodicity pitch mechanisms from the signals the CI-system provides even though the neural excitation pattern does not match the pattern expected in normal-hearing persons. A prerequisite is that the CI-system provides the necessary information in the stimulation pattern which at present is not the case because of the envelope extraction at relatively low frequencies. It may thus be interesting to study periodicity pitch in implanted children.

Attempts to improve the fine structure of the electrical stimulation pattern within the framework of the search for periodicity pitch in CI subjects are not only beneficial to pitch perception. The fine structure is also required to segregate multiple simultaneous sources (for instance in noisy background conditions or in polyphonic music) and for directional hearing.

8.5 Summary

In this thesis we developed a research platform that allows setting up arbitrary experiments with Nucleus CI subjects. We used this platform to assess the pitch mechanisms in CI subjects in a number of different studies. We showed that place pitch in CI subjects is related to the centroid of the excitation pattern and thus is closely related to spectral pitch in normal-hearing persons. We also showed that place pitch sensitivity appears to be related to the steepness of the edges of the excitation pattern. We showed that temporal pitch salience is highest when the 'in-phase' modulation depth over the channels is largest. We used this knowledge about the pitch mechanisms in CI subjects to build a new sound processing strategy (F0mod) designed to enhance pitch perception. With this new sound processing scheme pitch discrimination was three times better than with current sound processing strategies and also musical perception was improved.

The analogies that we found between the pitch mechanisms in CI subjects and the pitch mechanisms in normal-hearing persons, allow for creating an acoustic model of CI. This model provides normal-hearing listeners an impression of what CI subjects hear.

Although we obtained a significant improvement in pitch perception by CI subjects, there remains a large difference between their performance (even with the F0mod scheme) and the performance of normal-hearing persons. To overcome this gap future research should try to enable the periodicity pitch mechanism in CI subjects.

Bibliography

P. Allen, F. Wightman, D. Kistler, and T. Dolan. Frequency resolution in children. *Journal of Speech & Hearing Research,* 32(2):317-322, June 1989.

American National Standards Institute. *American National Standard Acoustical Terminology for physiological and psychological acoustics. ANSI S1.1-1994*. American Standards Association, New York, 1994.

J.N. Anantharaman, A.K. Krishnamurthy, and L.L. Feth. Intensity-weighted average of instantaneous frequency as a model for frequency discrimination. *Journal of the Acoustical Society of America,* 94(2 Pt 1):723-729, August 1993.

S.P. Bacon and N.F. Viemeister. Temporal-Modulation Transfer-Functions in Normal-Hearing and Hearing-Impaired Listeners. *Audiology,* 24(2):117-134, 1985.

J.G. Barry, P.J. Blamey, L.F. Martin, K.Y. Lee, T. Tang, Y.Y. Ming, and C.A. Van Hasselt. Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clinical Linguistic Phonetics,* 16(2):79-99, March 2002.

D. Baskent and R.V. Shannon. Speech recognition under conditions of frequency-place compression and expansion. *Journal of the Acoustical Society of America,* 113(4):2064-2076, 2003.

R.D. Battmer, I. Feldmeier, A. Kohlenberg, and T. Lenarz. Performance of the new Clarion speech processor 1.2 in quiet and in noise. *American Journal of Otology,* 18(6 Suppl):S144-S146, November 1997.

K. Bhatia, K.P. Gibbin, T.P. Nikolopoulos, and G.M. O'Donoghue. Surgical Complications and Their Management in a Series of 300 Consecutive Pediatric Cochlear Implantations. *Otology and Neurotology,* 25(5):730-739, September 2004.

R.C. Black and G.M. Clark. Differential electrical excitation of the auditory nerve. *Journal of the Acoustical Society of America,* 67(3):868-874, March 1980.

P.J. Blamey, R.C. Dowell, Y.C. Tong, A.M. Brown, S.M. Luscombe, and G.M. Clark. Speech processing studies using an acoustic model of a multiple-channel cochlear implant. *Journal of the Acoustical Society of America,* 76(1):104-110, July 1984a.

P.J. Blamey, R.C. Dowell, Y.C. Tong, and G.M. Clark. An acoustic model of a multiple-channel cochlear implant. *Journal of the Acoustical Society of America,* 76(1):97-103, July 1984b.

B. Boets, M. Van Ingelghem, J. Laneau, J. Wouters, A. van Wieringen, and P. Ghesquiere. Auditory temporal information processing in preschool children at risk of dyslexia : relations with phonological awareness and developing literacy skills. *Brain and Language,* submitted2005.

E.M. Burns and N.F. Viemeister. Non-Spectral Pitch. *Journal of the Acoustical Society of America,* 60(4):863-869, 1976.

P.A. Busby and G.M. Clark. Electrode discrimination by early-deafened cochlear implant patients. *Audiology,* 35(1):8-22, January 1996.

P.A. Busby and G.M. Clark. Electrode discrimination by early-deafened subjects using the cochlear limited multiple-electrode cochlear implant. *Ear and Hearing,* 21(4):291-304, August 2000a.

P.A. Busby and G.M. Clark. Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant. *Journal of the Acoustical Society of America,* 107(1):547-558, January 2000b.

P.A. Busby, Y.C. Tong, and G.M. Clark. Electrode position, repetition rate, and speech perception by early- and late-deafened cochlear implant patients. *Journal of the Acoustical Society of America,* 93(2):1058-1067, February 1993.

P.A. Cariani and B. Delgutte. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology,* 76(3):1698-1716, September 1996.

R.P. Carlyon and J.M. Deeks. Limitations on rate discrimination. *Journal of the Acoustical Society of America,* 112(3 Pt 1):1009-1025, September 2002.

R.P. Carlyon, L. Geurts, and J. Wouters. Detection of small across-channel timing differences by cochlear implantees. *Hearing Research,* 141(1-2):140-154, March 2000.

R.P. Carlyon, A. van Wieringen, C.J. Long, J.M. Deeks, and J. Wouters. Temporal pitch mechanisms in acoustic and electric hearing. *Journal of the Acoustical Society of America,* 112(2):621-633, August 2002.

M. Chatterjee. Modulation masking in cochlear implant listeners: envelope versus tonotopic components. *Journal of the Acoustical Society of America,* 113(4 Pt 1):2042-2053, April 2003.

H. Chen and F.G. Zeng. Frequency modulation detection in cochlear implant subjects. *Journal of the Acoustical Society of America,* 116(4):2269-2277, 2004.

Cochlear Ltd. Nucleus Implant Communicator (NIC) System Overview. Tech. Report N95291 Iss.1, Cochlear Ltd., December 2002.

Cochlear Ltd. Reliability update - end june 2003. *Nucleus Report,* 5, August 2003.

L.T. Cohen, P.A. Busby, and G.M. Clark. Cochlear implant place psychophysics. 2. Comparison of forward masking and pitch estimation data. *Audiology and Neurootology,* 1(5):278-292, September 1996a.

L.T. Cohen, P.A. Busby, L.A. Whitford, and G.M. Clark. Cochlear implant place psychophysics 1. Pitch estimation with deeply inserted electrodes. *Audiology and Neurootology,* 1(5):265-277, September 1996b.

L.T. Cohen, E. Saunders, and G.M. Clark. Psychophysics of a prototype peri-modiolar cochlear implant electrode array. *Hearing Research,* 155(1-2):63-81, May 2001.

M.A. Cohen, S. Grossberg, and L.L. Wyse. A spectral network model of pitch perception. *Journal of the Acoustical Society of America,* 98(2 Pt 1):862-879, August 1995.

L.M. Collins, T.A. Zwolan, and G.H. Wakefield. Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects. *Journal of the Acoustical Society of America,* 101(1):440-455, January 1997.

H.P. Dai, Q. Nguyen, G. Kidd, L.L. Feth, and D.M. Green. Phase independence of pitch produced by narrow-band sounds. *Journal of the Acoustical Society of America,* 100(4):2349-2351, 1996.

C.J. Darwin and R.P. Carlyon. *Auditory grouping*, chapter 11 in "Hearing" (Moore, B. C. J., Ed.), pages 387-424. Academic Press, 1995.

A. de Cheveigné. *Pitch perception models*, chapter 6 in "Pitch - Neural Coding and Perception" (Plack, C., Oxenham, A. J., Popper, A. N., and Fay, R. R., Eds.), Springer-Verlag, June 2005.

A. de Cheveigne and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America,* 111(4):1917-1930, April 2002.

J.M. Deeks and R.P. Carlyon. Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation. *Journal of the Acoustical Society of America,* 115(4):1736-1746, April 2004.

L. Demany and C. Semal. Pitch Versus Brightness of Timbre - Detecting Combined Shifts in Fundamental and Formant Frequency. *Music Perception,* 11(1):1-14, 1993.

A. Djourno and C. Eyries. Prothèse auditive par excitation électrique à distance du nerf sensoriel à l'aide d'un bobinage inclus à demeure. *Presse Medicale,* 63:65, 1957.

M.F. Dorman and P.C. Loizou. Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. *American Journal of Otology,* 18(6):S113-S114, 1997.

M.F. Dorman and P.C. Loizou. The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. *Ear and Hearing,* 19(2):162-166, 1998.

M.F. Dorman, P.C. Loizou, and D. Rainey. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America,* 102(4):2403-2411, 1997.

R. Drullman, J.M. Festen, and R. Plomp. Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America,* 95(2):1053-1064, February 1994.

D.K. Eddington. Speech discrimination in deaf subjects with cochlear implants. *Journal of the Acoustical Society of America,* 68(3):885-891, September 1980.

D.K. Eddington, W.H. Dobelle, D.E. Brackmann, M.G. Mladejovsky, and J.L. Parkin. Auditory prostheses research with multiple channel intracochlear stimulation in man. *Annals of Otology, Rhinology and Laryngology,* 87(6 Pt 2):1-39, November 1978.

A. Faulkner, S. Rosen, and C. Smith. Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants. *Journal of the Acoustical Society of America,* 108(4):1877-1887, 2000.

R. Fearn and J. Wolfe. Relative importance of rate and place: experiments using pitch scaling techniques with cochlear implants recipients. *Annals of Otology, Rhinology and Laryngology Supplement,* 185:51-53, December 2000.

R. Filipo, P. Mancini, D. Ballantyne, E. Bosco, and C. D'Elia. Short-term study of the effect of speech coding strategy on the auditory performance of pre- and post-lingually deafened adults implanted with the Clarion CII. *Acta Oto-Laryngologica,* 124(4):368-370, May 2004.

J.L. Flanagan and M.G. Saslow. Pitch discrimination for synthetic vowels. *Journal of the Acoustical Society of America,* 30:435-442, 1958.

C. Formby. Differential Sensitivity to Tonal Frequency and to the Rate of Amplitude-Modulation of Broad-Band Noise by Normally Hearing Listeners. *Journal of the Acoustical Society of America,* 78(1):70-77, 1985.

C. Formby and K. Muir. Modulation and gap detection for broadband and filtered noise signals. *Journal of the Acoustical Society of America,* 84(2):545-550, August 1988.

L.M. Friesen, R.V. Shannon, D. Baskent, and X. Wang. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America,* 110(2):1150-1163, 2001.

Q.J. Fu, S. Chinchilla, and J.J. Galvin. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology,* 5(3):253-260, September 2004.

Q.J. Fu and G. Nogaki. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology,* in press2005.

Q.J. Fu and R.V. Shannon. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America,* 105(3):1889-1900, 1999.

Q.J. Fu and R.V. Shannon. Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners. *Journal of the Acoustical Society of America,* 107(1):589-597, January 2000.

Q.J. Fu, R.V. Shannon, and J.J. Galvin, III. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of the Acoustical Society of America,* 112(4):1664-1674, October 2002.

Q.J. Fu, F.G. Zeng, R.V. Shannon, and S.D. Soli. Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America,* 104(1):505-510, 1998.

J.P. Gagne and P.M. Zurek. Resonance-frequency discrimination. *Journal of the Acoustical Society of America,* 83(6):2293-2299, June 1988.

L. Geurts and J. Wouters. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. *Journal of the Acoustical Society of America,* 105(4):2476-2484, April 1999.

L. Geurts and J. Wouters. A concept for a research tool for experiments with cochlear implant users. *Journal of the Acoustical Society of America,* 108(6):2949-2956, December 2000.

L. Geurts and J. Wouters. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *Journal of the Acoustical Society of America,* 109(2):713-726, February 2001.

L. Geurts and J. Wouters. Better place-coding of the fundamental frequency in cochlear implants. *Journal of the Acoustical Society of America,* 115(2):844-852, February 2004.

K. Gfeller, A. Christ, J.F. Knutson, S. Witt, K.T. Murray, and R.S. Tyler. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology,* 11(7):390-406, July 2000.

K. Gfeller and C. Lansing. Musical Perception of Cochlear Implant Users As Measured by the Primary Measures of Music Audiation - An Item Analysis. *Journal of Music Therapy,* 29(1):18-39, 1992.

K. Gfeller and C.R. Lansing. Melodic, Rhythmic, and Timbral Perception of Adult Cochlear Implant Users. *Journal of Speech and Hearing Research,* 34(4):916-920, August 1991.

K. Gfeller, C. Turner, M. Mehr, G. Woodworth, R. Fearn, J. Knutson, S. Witt, and J. Stordahl. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International,* 3(1):29-53, 2002.

K. Gfeller, G. Woodworth, D.A. Robin, S. Witt, and J.F. Knutson. Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and Hearing,* 18(3):252-260, June 1997.

B.R. Glasberg and B.C. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research,* 47(1-2):103-138, August 1990.

H. Gockel, R.P. Carlyon, and C.J. Plack. Across-frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms. *Journal of the Acoustical Society of America,* 116(2):1092-1104, August 2004.

J.L. Goldstein. An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America,* 54(6):1496-1516, December 1973.

K.W. Grant, V. Summers, and M.R. Leek. Modulation rate detection and discrimination by normal hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America,* 104(2):1051-1060, August 1998.

D.M. Green and J.A. Swets. *Signal detection theory and psychophysics*. Peninsula Publishing, Los Altos, California USA, 1988.

T. Green, A. Faulkner, and S. Rosen. Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *Journal of the Acoustical Society of America,* 112(5 Pt 1):2155-2164, November 2002.

T. Green, A. Faulkner, and S. Rosen. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. *Journal of the Acoustical Society of America,* 116(4):2289-2297, 2004.

D.D. Greenwood. A cochlear frequency-position function for several species--29 years later. *Journal of the Acoustical Society of America,* 87(6):2592-2605, June 1990.

J.M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America,* 61(5):1270-1277, May 1977.

T.D. Griffiths, C. Buchel, R.S. Frackowiak, and R.D. Patterson. Analysis of temporal structure in sound by the human brain. *Nature Neuroscience,* 1(5):422-427, September 1998.

T.D. Griffiths, S. Uppenkamp, I. Johnsrude, O. Josephs, and R.D. Patterson. Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience,* 4(6):633-637, June 2001.

N. Grimault, C. Micheyl, R.P. Carlyon, and L. Collet. Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Perception and Psychophysics,* 64(2):189-197, February 2002.

E.R. Hafter and S.C. Carrier. Binaural interaction in low-frequency stimuli: the inability to trade time and intensity completely. *Journal of the Acoustical Society of America,* 51(6):1852-1862, June 1972.

T.E. Hanna. Discrimination and identification of modulation rate using a noise carrier. *Journal of the Acoustical Society of America,* 91(4 Pt 1):2122-2128, April 1992.

R. Hartmann and A. Kral. *Central responses to electrical stimulation*, chapter 6 in "Cochlear implants - Auditory prostheses and electric hearing" (Zeng, F. G., Popper, A. N., and Fay, R. R., Eds.), pages 213-285. Springer Verlag, 2004.

B.A. Henry, C.M. McKay, H.J. McDermott, and G.M. Clark. The relationship between speech perception and electrode discrimination in cochlear implantees. *Journal of the Acoustical Society of America,* 108(3 Pt 1):1269-1280, September 2000.

D. Hilbert. *Grundzüge einer allgemeinen theorie der linearen integralgleichungen*. Teubner, Leipzig, Germany, 1912.

A.J.M. Houtsma and J. Smurzynski. Pitch Identification and Discrimination for Complex Tones with Many Harmonics. *Journal of the Acoustical Society of America,* 87(1):304-310, January 1990.

C.H. Huang, J.B. Laflen, and T.M. Talavage. Generation of acoustic stimuli from a modeled cochlear implant-induced neural activation pattern. *International Congress Series,* 1273:52-55, 2004.

C.J. James, P.J. Blamey, L. Martin, B. Swanson, Y. Just, and D. Macfarlane. Adaptive dynamic range optimization for cochlear implants: a preliminary study. *Ear and Hearing,* 23(1 Suppl):49S-58S, February 2002.

P.A. Jones, H.J. McDermott, P.M. Seligman, and J.B. Millar. Coding of voice source information in the Nucleus cochlear implant system. *Annals of Otology, Rhinology and Laryngology Supplement,* 166:363-365, September 1995.

R.C. Jones, S.S. Stevens, and M.H. Lurie. Three mechanisms of hearing by electrical stimulation. *Journal of the Acoustical Society of America,* 12:281-290, 1940.

C. Kaernbach and L. Demany. Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustical Society of America,* 104(4):2298-2306, October 1998.

D.H. Klatt. Software for a cascade-parallel formant synthesizer. *Journal of the Acoustical Society of America,* 67(3):971-995, 1980.

D.B. Koch, M.J. Osberger, P. Segel, and D. Kessler. HiResolution and conventional sound processing in the HiResolution bionic ear: using appropriate outcome measures to assess speech recognition ability. *Audiology and Neurootology,* 9(4):214-223, July 2004.

A. Kohlrausch, R. Fassel, and T. Dau. The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *Journal of the Acoustical Society of America,* 108(2):723-734, August 2000.

Y.Y. Kong, R. Cruz, J.A. Jones, and F.G. Zeng. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing,* 25(2):173-185, April 2004.

Y.Y. Kong, G.S. Stickney, and F.G. Zeng. Speech and melody recognition in binaurally combined acoustic and electric hearing. *Journal of the Acoustical Society of America,* 117(3):1351-1361, March 2005.

K. Krumbholz, R.D. Patterson, A. Seither-Preisler, C. Lammertmann, and B. Lutkenhoner. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex,* 13(7):765-772, July 2003.

N. Lan, K.B. Nie, S.K. Gao, and F.G. Zeng. A novel speech-processing strategy incorporating tonal information for cochlear implants. *IEEE Transactions on Biomedical Engineering,* 51(5):752-760, May 2004.

J. Laneau, M. Moonen, and J. Wouters. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *Journal of the Acoustical Society of America,* 116(6):3606-3619, December 2004.

J. Laneau, M. Moonen, and J. Wouters. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *Journal of the Acoustical Society of America,* submitted2005.

J. Laneau and J. Wouters. Multi-channel place pitch sensitivity in cochlear implant recipients. *Journal of the Association for Research in Otolaryngology,* 5(3):285-294, 2004.

M.C. Leal, Y.J. Shin, M.L. Laborde, M.N. Calmels, S. Verges, S. Lugardon, S. Andrieu, O. Deguine, and B. Fraysse. Music perception in adult cochlear implant recipients. *Acta Oto-Laryngologica,* 123(7):826-835, September 2003.

H. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America,* 49:467-477, September 1971.

J.C.R. Licklider. A duplex theory of pitch perception. *Experientia,* 7:128-134, 1951.

J.C.R. Licklider. Periodicity pitch and place pitch. *Journal of the Acoustical Society of America,* 26:945, 1954.

P.C. Loizou, M. Dorman, O. Poroy, and T. Spahr. Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution. *Journal of the Acoustical Society of America,* 108(5):2377-2387, 2000.

X. Luo and Q.J. Fu. Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *Journal of the Acoustical Society of America,* 116(6):3659-3667, December 2004.

J. Lyzenga and J.W. Horst. Frequency Discrimination of Band-Limited Harmonic Complexes Related to Vowel Formants. *Journal of the Acoustical Society of America,* 98(4):1943-1955, October 1995.

J. Lyzenga and J.W. Horst. Frequency discrimination of stylized synthetic vowels with a single formant. *Journal of the Acoustical Society of America,* 102(3):1755-1767, September 1997.

O. Macherey. *Evaluation of a method to improve perception of voice pitch in users of CIS cochlear implants*. Master thesis, University College London, 2003.

N.A. MacMillan and C.D. Creelman. *Detection theory: A user's guide*. Cambridge University Press, Cambridge, 1991.

J.D. Markel. The SIFT algorithm for fundamental frequency estimation. *IEEE Transactions on audio and electroacoustics,* 20(5):367-377, December 1972.

H.J. McDermott and C.M. McKay. Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. *Journal of the Acoustical Society of America,* 96(1):155-162, July 1994.

H.J. McDermott and C.M. McKay. Musical pitch perception with electrical stimulation of the cochlea. *Journal of the Acoustical Society of America,* 101(3):1622-1631, March 1997.

H.J. McDermott, C.M. McKay, L.M. Richardson, and K.R. Henshall. Application of loudness models to sound processing for cochlear implants. *Journal of the Acoustical Society of America,* 114(4 Pt 1):2190-2197, October 2003.

C.M. McKay and R.P. Carlyon. Dual temporal pitch percepts from acoustic and electric amplitude- modulated pulse trains. *Journal of the Acoustical Society of America,* 105(1):347-357, January 1999.

C.M. McKay and H.J. McDermott. The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites. *Journal of the Acoustical Society of America,* 100(2 Pt 1):1081-1092, August 1996.

C.M. McKay, H.J. McDermott, and R.P. Carlyon. Place and temporal cues in pitch perception: are they truly independent? *Acoustic Research Letters Online,* 1(1):25-30, July 2000.

C.M. McKay, H.J. McDermott, and G.M. Clark. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *Journal of the Acoustical Society of America,* 96(5 Pt 1):2664-2673, November 1994.

C.M. McKay, H.J. McDermott, and G.M. Clark. The perceptual dimensions of single-electrode and nonsimultaneous dual- electrode stimuli in cochlear implantees. *Journal of the Acoustical Society of America,* 99(2):1079-1090, February 1996.

C.M. McKay, A. O'Brien, and C.J. James. Effect of current level on electrode discrimination in electrical stimulation. *Hearing Research,* 136(1-2):159-164, October 1999.

R. Meddis and M.J. Hewitt. Virtual Pitch and Phase Sensitivity of A Computer-Model of the Auditory Periphery .1. Pitch Identification. *Journal of the Acoustical Society of America,* 89(6):2866-2882, June 1991.

B.C.J. Moore and R.P. Carlyon. *Perception of pitch by people with cochlear hearing loss and by cochlear implant users*, chapter 7 in "Pitch - Neural coding and perception" (Plack, C., Oxenham, A. J., Fay, R. R., and Popper, A. N., Eds.), Springer-Verlag, June 2005.

B.C.J. Moore and S.M. Rosen. Tune recognition with reduced pitch and interval information. *Quarterly Journal for Experimental Psychology,* 31(2):229-240, May 1979.

D.J. Morris and B.E. Pfingst. Effects of electrode configuration and stimulus level on rate and level discrimination with cochlear implants. *Journal of the Association for Research in Otolaryngology,* 1(3):211-223, November 2000.

T. Nakatani and T. Irino. Robust and accurate fundamental frequency estimation based on dominant harmonic components. *Journal of the Acoustical Society of America,* 116(6):3690-3700, December 2004.

D.A. Nelson, G.S. Donaldson, and H.A. Kreft. Spatial selectivity, neurotopicity and tonotopicity in cochlear implant listeners. In *Abstracts presented at the 2003 Conference on Implantable Auditory Prostheses*, page 38, Asilomar, CA, USA, August 2003a.

D.A. Nelson, D.J. Van Tasell, A.C. Schroder, S. Soli, and S. Levine. Electrode ranking of "place pitch" and speech recognition in electrical hearing. *Journal of the Acoustical Society of America,* 98(4):1987-1999, October 1995.

P.B. Nelson, S.H. Jin, A.E. Carney, and D.A. Nelson. Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America,* 113(2):961-968, 2003b.

J.K. Niparko. *Cochlear Implants: clinical applications*, chapter 3 in "Cochlear implants: auditory prostheses and electric hearing" (Zeng, F. G., Popper, A. N., and Fay, R. R., Eds.), pages 53-100. Springer-Verlag, 2004.

K. Ohgushi. On the role of spatial and temporal cues in the perception of the pitch of complex tones. *Journal of the Acoustical Society of America,* 64(3):764-771, September 1978.

A.V. Oppenheim and R.W. Schafer. *Discrete-time signal processing*. Prentice Hall, Inc., Upper Saddle River, New Jersey, 1999.

M.J. Osberger and L. Fisher. New directions in speech processing: patient performance with simultaneous analog stimulation. *The Annals of otology, rhinology & laryngology Supplement,* 185:70-73, December 2000.

A.J. Oxenham, J.G.W. Bernstein, and H. Penagos. Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences of the United States of America,* 101(5):1421-1425, February 2004.

R.D. Patterson, M.H. Allerhand, and C. Giguere. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *Journal of the Acoustical Society of America,* 98(4):1890-1894, October 1995.

R.D. Patterson, D. Johnson-Davies, and R. Milroy. Amplitude-Modulated Noise - Detection of Modulation Versus Detection of Modulation Rate. *Journal of the Acoustical Society of America,* 63(6):1904-1911, 1978.

R.D. Patterson, S. Uppenkamp, I.S. Johnsrude, and T.D. Griffiths. The processing of temporal pitch and melody information in auditory cortex. *Neuron,* 36(4):767-776, November 2002.

H. Penagos, J.R. Melcher, and A.J. Oxenham. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience,* 24(30):6810-6815, July 2004.

B.E. Pfingst, K.H. Franck, L. Xu, E.M. Bauer, and T.A. Zwolan. Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses. *Journal of the Association for Research in Otolaryngology,* 2(2):87-103, June 2001.

B.E. Pfingst, L.A. Holloway, T.A. Zwolan, and L.M. Collins. Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants. *Hearing Research,* 134(1-2):105-115, August 1999.

J.O. Pickles. *An introduction to the physiology of hearing*. Academic Press Limited, London, 1988.

S. Pijl. Labeling of musical interval size by cochlear implant patients and normally hearing subjects. *Ear and Hearing,* 18(5):364-372, October 1997.

S. Pijl and D.W. Schwarz. Intonation of musical intervals by musical intervals by deaf subjects stimulated with single bipolar cochlear implant electrodes. *Hearing Research,* 89(1-2):203-211, September 1995a.

S. Pijl and D.W. Schwarz. Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *Journal of the Acoustical Society of America,* 98(2 Pt 1):886-895, August 1995b.

C. Plack and A.J. Oxenham. *Overview: The present anf future of Pitch*, chapter 1 in "Pitch - Neural coding and perception" (Plack, C., Oxenham, A. J., Popper, A. N., and Fay, R. R., Eds.), Springer-Verlag, June 2005a.

C. Plack and A.J. Oxenham. *The psychophysics of Pitch*, chapter 2 in "Pitch - Neural coding and perception" (Plack, C., Oxenham, A. J., Popper, A. N., and Fay, R. R., Eds.), Springer-Verlag, June 2005b.

R. Plomp and H.J.M. Steeneken. Pitch versus timbre. In *Proceedings of the seventh international conference on acoustics*, pages 377-380, Budapest, Hungaria, 1971.

M.K. Qin and A.J. Oxenham. F0 discriminability and utility with acoustic simulation of cochlear implant signal processing. *Ear and Hearing,* submitted2005.

L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, and C.A. McGonegal. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on acoustics, speech, and signal processing,* 24(5):399-418, October 1976.

R. Renken, J.E. Wiersinga-Post, S. Tomaskovic, and H. Duifhuis. Dominance of missing fundamental versus spectrally cued pitch: individual differences for complex tones with unresolved harmonics. *Journal of the Acoustical Society of America,* 115(5 Pt 1):2257-2263, May 2004.

J.T. Rubinstein, B.S. Wilson, C.C. Finley, and P.J. Abbas. Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing Research,* 127(1-2):108-118, January 1999.

E. Schulz and M. Kerber. *Music perception with the MED-EL implants*, chapter in "Advances in cochlear implants" (Hochmair-Desoyer, I. J. and Hochmair, E. S., Eds.), pages 326-332. Manz, 1993.

A. Sek and B.C. Moore. Discrimination of frequency steps linked by glides of various durations. *Journal of the Acoustical Society of America,* 106(1):351-359, July 1999.

P. Seligman and H. McDermott. Architecture of the Spectra 22 speech processor. *The Annals of otology, rhinology & laryngology Supplement,* 166:139-141, September 1995.

T.M. Shackleton and R.P. Carlyon. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *Journal of the Acoustical Society of America,* 95(6):3529-3540, June 1994.

S. Shamma and D. Klein. The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America,* 107(5 Pt 1):2631-2644, May 2000.

R.V. Shannon. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hearing Research,* 11(2):157-189, August 1983a.

R.V. Shannon. Multichannel electrical stimulation of the auditory nerve in man. II. Channel Interaction. *Hearing Research,* 12(1):1-16, 1983b.

R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. *Science,* 270(5234):303-304, October 1995.

R.V. Shannon, F.G. Zeng, and J. Wygonski. Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America,* 104(4):2467-2476, 1998.

R.K. Shepherd and E. Javel. Electrical stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status. *Hearing Research,* 108(1-2):112-144, June 1997.

M.W. Skinner, P.L. Arndt, and S.J. Staller. Nucleus 24 advanced encoder conversion study: performance versus preference. *Ear and Hearing,* 23(1 Suppl):2S-17S, February 2002a.

M.W. Skinner, L.K. Holden, L.A. Whitford, K.L. Plant, C. Psarros, and T.A. Holden. Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear and Hearing,* 23(3):207-223, June 2002b.

M. Slaney. An efficient implementation of the Patterson-Holdsworth auditory filter bank. Tech. Report 45, Apple Computer, 1993.

Z.M. Smith, B. Delgutte, and A.J. Oxenham. Chimaeric sounds reveal dichotomies in auditory perception. *Nature,* 416(6876):87-90, March 2002.

G.F. Smoorenburg. Pitch perception of two-frequency stimuli. *Journal of the Acoustical Society of America,* 48(4):924-942, October 1970.

R.L. Snyder, J.A. Bierer, and J.C. Middlebrooks. Topographic spread of inferior colliculus activation in response to acoustic and intracochlear electric stimulation. *Journal of the Association for Research in Otolaryngology,* 5(3):305-322, September 2004.

D.R. Soderquist and R.D. Shilling. Forward Masking in Young-Children - Rocketship Psychophysics. *Current Psychology-Research & Reviews,* 11(2):145-156, 1992.

S.S. Stevens. On hearing by electrical stimulation. *Journal of the Acoustical Society of America,* 8:191-195, 1937.

S.S. Stevens and R.C. Jones. The mechanism of hearing by electrical stimulation. *Journal of the Acoustical Society of America,* 10:261-269, 1939.

L.J. Stover and L.L. Feth. Pitch of narrow-band signals. *Journal of the Acoustical Society of America,* 73(5):1701-1707, May 1983.

M.A. Svirsky, A.M. Robbins, K.I. Kirk, D.B. Pisoni, and R.T. Miyamoto. Language development in profoundly deaf children with cochlear implants. *Psychological Science,* 11(2):153-158, March 2000.

B.G. Tabachnick and L.S. Fidell. *Using multivariate statistics*. HarperCollins College Publishers, Northridge, 1996.

E. Terhardt. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America,* 55(5):1061-1069, May 1974.

Y.C. Tong, P.J. Blamey, R.C. Dowell, and G.M. Clark. Psychophysical Studies Evaluating the Feasibility of A Speech Processing Strategy for A Multiple-Channel Cochlear Implant. *Journal of the Acoustical Society of America,* 74(1):73-80, 1983.

Y.C. Tong and G.M. Clark. Absolute identification of electric pulse rates and electrode positions by cochlear implant patients. *Journal of the Acoustical Society of America,* 77(5):1881-1888, May 1985.

B. Townshend, N. Cotter, D. Van Compernolle, and R.L. White. Pitch perception by cochlear implant subjects. *Journal of the Acoustical Society of America,* 82(1):106-115, July 1987.

M. Van Ingelghem, A. van Wieringen, J. Wouters, E. Vandenbussche, P. Onghena, and P. Ghesquiere. Psychophysical evidence for a general temporal processing deficit in children with dyslexia. *Neuroreport,* 12(16):3603-3607, November 2001.

A. van Wieringen, R.P. Carlyon, J. Laneau, and J. Wouters. Effects of waveform shape on human sensitivity to electrical stimulation of the inner ear. *Hearing Research,* 200:73-86, February 2005.

A. van Wieringen, R.P. Carlyon, C.J. Long, and J. Wouters. Pitch of amplitude-modulated irregular-rate stimuli in acoustic and electric hearing. *Journal of the Acoustical Society of America,* 114(3):1516-1528, September 2003.

A. van Wieringen and J. Wouters. Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees. *Journal of the Acoustical Society of America,* 106(4 Pt 1):1925-1939, October 1999a.

A. van Wieringen and J. Wouters. Natural vowel and consonant recognition by Laura cochlear implantees. *Ear and Hearing,* 20(2):89-103, April 1999b.

A. van Wieringen and J. Wouters. Comparison of procedures to determine electrical stimulation thresholds in cochlear implant users. *Ear and Hearing,* 22(6):528-538, December 2001.

A.E. Vandali. Emphasis of short-duration acoustic speech cues for cochlear implant users. *Journal of the Acoustical Society of America,* 109(5):2049-2061, May 2001.

A.E. Vandali, L.A. Whitford, K.L. Plant, and G.M. Clark. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and Hearing,* 21(6):608-624, December 2000.

N.J. Versfeld. Discrimination of changes in the spectral shape of noise bands. *Journal of the Acoustical Society of America,* 102(4):2264-2275, 1997.

J. Ville. Théorie et applications de la notion de signal analytique. *Cables Transmission,* 2:61-74, 1948.

J.D. Warren, S. Uppenkamp, R.D. Patterson, and T.D. Griffiths. Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America,* 100(17):10038-10042, August 2003.

F. Wightman, P. Allen, T. Dolan, D. Kistler, and D. Jamieson. Temporal resolution in children. *Child Development,* 60(3):611-624, June 1989.

B.S. Wilson. The future of cochlear implants. *British Journal of Audiology,* 31(4):205-225, August 1997.

B.S. Wilson. *Strategies for representing speech with cochlear implants*, chapter 7 in "Cochlear implants - principles and practices" (Niparko, J. K., Ed.), pages 129-170. Lippincott Williams and Wilkins, 2000.

B.S. Wilson. *Engineering design of cochlear implants*, chapter 2 in "Cochlear implants: auditory prostheses and electric hearing" (Zeng, F. G., Popper, A. N., and Fay, R. R., Eds.), pages 14-52. Springer-Verlag, 2004.

B.S. Wilson, C.C. Finley, D.T. Lawson, R.D. Wolford, D.K. Eddington, and W.M. Rabinowitz. Better speech recognition with cochlear implants. *Nature,* 352(6332):236-238, July 1991.

J. Wouters, W. Damman, and A.J. Bosman. Vlaamse opname van woordenlijsten voor spraakaudiometrie. *Logopedie,* 7(6):28-33, 1994.

J. Wouters, L. Geurts, S. Peeters, J. Vanden Berghe, and A. van Wieringen. Developments in speech processing for cochlear implants. *Acta Otorhinolaryngologica Belgica,* 52(2):129-132, 1998.

L. Xu, Y.J. Tsai, and B.E. Pfingst. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America,* 112(1):247-258, 2002.

F.G. Zeng. Temporal pitch in electric hearing. *Hearing Research,* 174(1-2):101-106, December 2002.

F.G. Zeng. *Auditory Prostheses: Past, Present, and Future*, chapter 1 in "Cochlear Implants: Auditory prostheses and electric hearing" (Zeng, F. G., Popper, A. N., and Fay, R. R., Eds.), pages 1-13. Springer-Verlag, 2004.

F.G. Zeng and R.V. Shannon. Loudness-Coding Mechanisms Inferred from Electric-Stimulation of the Human Auditory-System. *Science,* 264(5158):564-566, April 1994.

M. Ziese, A. Stutzel, H. von Specht, K. Begall, B. Freigang, S. Sroka, and P. Nopp. Speech understanding with the CIS and the n-of-m strategy in the MED-EL COMBI 40+ system. *Journal for Oto-Rhino-Laryngology and its Related Specialties,* 62(6):321-329, November 2000.

E. Zwicker. *Masking and psychological excitation as consequences of the ear's frequency analysis*, chapter in "Frequency Analysis and Periodicity Detection in Hearing" (Plomp, R. and Smoorenburg, G. F., Eds.), pages 376-396. Sijthoff, 1970.

E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer-Verlag, Berlin, 1999.

T.A. Zwolan, L.M. Collins, and G.H. Wakefield. Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *Journal of the Acoustical Society of America,* 102(6):3673-3685, December 1997.

T.A. Zwolan, P.R. Kileny, C. Ashbaugh, and S.A. Telian. Patient performance with the Cochlear Corporation "20+2" implant: Bipolar versus monopolar activation. *American Journal of Otology,* 17(5):717-723, September 1996.

List of publications

Publications in international journals

J. Wouters, L. Geurts, J. Laneau, H. Luts, J.B. Maj, M. Moonen, L. Royackers, A. Spriet, T. Van den Bogaert, and A. van Wieringen. Digital hearing aids and future directions for hearing aids. *Acta Oto-Rhino-Laryngologica Belgica,* 56:357-361, 2003.

J. Laneau and J. Wouters. Multi-channel place pitch sensitivity in cochlear implant recipients. *Journal of the Association for Research in Otolaryngology,* 5(3):285-294, September 2004.

J. Laneau, M. Moonen, and J. Wouters. Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. *Journal of the Acoustical Society of America,* 116(6):3606-3619, December 2004.

A. van Wieringen, R.P. Carlyon, J. Laneau, and J. Wouters. Effects of waveform shape on human sensitivity to electrical stimulation of the inner ear. *Hearing Research,* 200:73-86, February 2005.

J. Laneau, B. Boets, M. Moonen, A. van Wieringen, and J. Wouters. A flexible auditory research platform using acoustic or electric stimuli for adults and young children. *Journal of Neuroscience Methods,* 142:131-136, March 2005.

J. Laneau, M. Moonen, and J. Wouters. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *Journal of the Acoustical Society of America,* In review 2005.

J. Laneau, M. Moonen, and J. Wouters. Improved music perception with explicit pitch coding in cochlear implants. *Audiology and Neuro-Otology,* In review 2005.

Conference proceedings and other publications

J. Laneau and J. Wouters. Fractal effects and models of electrically stimulated fibers. In proceedings of the *World Congress on Neuroinformatics*, pages 170-178, Vienna, Austria, September 2001.

J. Laneau. Building APEX: a research patform for Nucleus CI22 and CI24 cochlear implants. *Nucleus Report,* August/September:4, 2003.

J. Laneau and J. Wouters. Pitch perception in cochlear implants with different filter bank designs. In proceedings of the *IEEE Engineering in Medicine and Biology Conference (EMBC)*, pages 1995-1998 Vol. 3, Cancun, Mexico, September 2003.

J. Laneau and J. Wouters. Music and tonal language processing with NICstream. In proceedings of the *European Nucleus Outcomes Seminar*, Sardinia, Italy, September 2003.

Contributions to conferences

J. Wouters, H. Luts, J. Laneau, and L. Mens. AMFR procedures for frequency specific objective threshold estimation. Presented at the *Conference of the International Collegium for Rehabilitative Audiology (ICRA),* Cardiff, UK, May 2001.

J. Wouters, L. Geurts, J. Laneau, H. Luts, J.B. Maj, M. Moonen, A. Spriet, and A. van Wieringen. More signal processing for better speech perception? Presented at the *Conference on Implantable Auditory Protheses (CIAP)*,Asilomar, CA, USA, August 2001.

J. Laneau and J. Wouters. Fractal effects and models of electrically stimulated fibers. Presented at the *World Congress on Neuroinformatics,* Vienna, Austria, September 2001.

J. Laneau and J. Wouters. Extension of APEX to Nucleus Implants - a (too) early experimence with NICstream. Presented at the *NIC Workshop,* Alpbach, Austria, January 2002.

J. Laneau and J. Wouters. Filterbanks for improved pitch coding in cochlear implants. *Belgian Day on Biomedical Engineering,* Brussels, Belgium, October 2002.

J. Laneau and J. Wouters. Filterbanks for improved pitch coding in cochlear implants. *PhD Symposium,* Leuven, Belgium, December 2002.

J. Laneau, L. Geurts, and J. Wouters. APEX 2.0. Presented at the *NIC Workshop,* Alpbach, Austria, January 2003.

J. Laneau, B. Boets, L. Geurts, M. Moonen, A. van Wieringen, and J. Wouters. APEX 2.1 - A user friendly and flexible research platform for auditory research using either acoustic stimuli or electric stimuli. *Conference on Implantable Auditory Protheses (CIAP),* Asilomar, USA, August 2003.

J. Laneau, M. Moonen, and J. Wouters. Noise band vocoders for cochlear implant pitch research in normal-hearing subjects. *Conference on Implantable Auditory Protheses (CIAP),* Asilomar, USA, August 2003.

J. Laneau, M. Moonen, and J. Wouters. The effect of filter banks on the pitch discrimination in cochlear implants. *Conference on Implantable Auditory Protheses (CIAP),* Asilomar, USA, August 2003.

J. Wouters, A. van Wieringen, R.P. Carlyon, and J. Laneau. Effects of filter bank structures and stimulus waveforms in cochlear implants. *Conference on Implantable Auditory Protheses (CIAP),* Asilomar, USA, August 2003.

J. Wouters and J. Laneau. Signal processing for improved sound perception in cochlear implants. Presented at the *IEEE Engineering in Medicine and Biology Conference (EMBC),* Cancun, Mexico, September 2003.

J. Laneau and J. Wouters. Music and tonal language processing using NICstream. Presented at the *European Nucleus Outcomes Seminar,* Cagliari, Sardinia, September 2003.

J. Laneau and J. Wouters. Music and tonal language perception in cochlear implants. Presented at the *DSP Valley Annual Research and Technology Symposium (DARTS),* Leuven, Belgium, October 2003.

J. Laneau and J. Wouters. Pitch perception in cochlear implant recipients. Presented at the *Belgian Society of Audiology (BSA) meeting,* Liège, Belgium, December 2003.

J. Laneau, M. Moonen, and J. Wouters. Pitch perception in cochlear implants. *PhD Symposium,* Brussels, Belgium, December 2003.

B. Van Dun, J. Wouters, M. Moonen, A. Vergult, J. Welkenhuyzen, J. Laneau, and H. Luts. Improvement of response detection with SSEP after comparison of several statistical tests. Presented at the *Belgian Society of Audiology (BSA) meeting,* Liège, Belgium, December 2003.

B. Boets, M. Van Ingelghem, J. Laneau, J. Wouters, A. van Wieringen, and P. Ghesquière. Auditory temporal information processing in preschool children at risk of dyslexia : relations with phonological awareness and developing literacy skills.

Presented at the *International Conference of British Dyslexia Association (BDA),* Warwick, UK, March 2004.

M. Büchler and J. Laneau. An Acoustic Model Using the Nucleus Matlab Toolbox. Presented at the *NIC Workshop,* Mechelen, Belgium, April 2004.

J. Laneau, B. Boets, O. Macherey, A. van Wieringen, and J. Wouters. Further Advances with APEX. Presented at the *NIC Workshop,* Mechelen, Belgium, April 2004.

J. Laneau, M. Moonen, and J. Wouters. Improved music perception using a new sound processor based upon explicit pitch coding. *European Symposium on Paediatric Cochlear Implantation,* Geneva, Switzerland, May 2004.

J. Laneau, J. Wouters, and M. Moonen. Sound processing for improved pitch perception in cochlear implantees. Presented at the *Mid-winter Meeting of the Association for Research in Otolaryngology (ARO),* New-Orleans, USA, February 2005.

L. Van Deun, A. Spriet, S. Doclo, K. Eftaxiadis, J. Laneau, J.-B. Maj, M. Moonen, B. Van Dijk, A. van Wieringen, J. Wouters. Benefit of a two-microphone adaptive beamformer in the new Nucleus cochlear implant system. Presented at the *International Collegium for Rehabilitative Audiology (ICRA)*, Gainesville, USA, March 2005.

J. Wouters, A. Spriet, L. Van Deun, S. Doclo, K. Eftaxiadis, J. Laneau, M. Moonen, B. Van Dijk, A. van Wieringen. Enhanced speech understanding with a two-microphone adaptive beamformer in the new Nucleus cochlear implant system. Presented at the *10th Symposium on Cochlear Implants in Children*, Dallas, USA, March 2005.

Curriculum vitae

Johan Laneau was born in Ukkel, Belgium, on the 10th March 1977. During his study of Applied Sciences at the Katholieke Universiteit Leuven, Belgium, he visited the Escuela Superior de Ingenieros in Sevilla, Spain for one year in 1998-1999. In 2000, he received the degree of Master in Electrical Engineering from the Katholieke Universiteit Leuven.

In October 2000, he started as a Ph. D. student at the Laboratory for Experimental ORL (Faculty of Medicine) of the Katholieke Universiteit Leuven. His supervisors are Prof. Jan Wouters and Prof. Marc Moonen. His research interests lay in the area of signal processing and software development for sound applications in general, and for hearing aids and cochlear implants in particular.

Johan Laneau is married to Lotte Loones and currently has one daughter Engelien.