

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)
In samenwerking met:
FACULTEIT GENEESKUNDE
DEPARTEMENT NEUROWETENSCHAPPEN
O&N2, Herestraat 49 bus 721, 3000 Leuven

Perception of binaural localization cues with combined electric and acoustic hearing

Jury:
Prof. Dr. Ir. A. Haegemans, voorzitter
Prof. Dr. J. Wouters, promotor
Prof. Dr. Ir. M. Moonen, promotor
Prof. Dr. Ir. D. Van Compernolle, assessor
Prof. Dr. Ir. H. Van Hamme, assessor
Dr. Ir. J.P.L. Brokx
(AZ Maastricht, The Netherlands)
Prof. Dr. A. Kohlrausch
(TU Eindhoven, The Netherlands)
Prof. Dr. B.C.J. Moore
(University of Cambridge, UK)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Tom Francart

U.D.C. 681.3*J3-534.7 November 2008

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

Alle rechten voorbehouden. Niets uit deze uitgave mag vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch 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/2008/7515/88

ISBN 978-90-5682-978-0

Voor Annelies

Voorwoord

Ruim vier jaar geleden deed ik een verkennende ronde op het departement elektrotechniek, op zoek naar een interessant onderwerp en leuk lab voor een doctoraat. Bijna was het cryptografie geworden, maar toen ik prof. Marc Moonen liet weten dat ik bij voorkeur de focus niet op wiskunde wou leggen, verwees hij me door naar een zekere prof. Jan Wouters die zijn thuisbasis had in het exotische St-Rafaël ziekenhuis, diep in de binnenstad van Leuven. Jan bleek uitzonderlijk gemotiveerd en had tal van interessante onderwerpen in de aanbieding. Ook van zijn toenmalige medewerkers hoorde ik niets dan goeds, dus de beslissing was snel genomen.

Wat me erg aantrok aan het onderzoek bij ExpORL, is de combinatie van een belangrijke technische component met een medische/menselijke component. Niet alleen was dat medische aspect een mogelijkheid om mijn horizon te verruimen, maar ook geeft het de relevantie aan van het technische werk. Hoewel ik dikwijls heb gesakkerd bij mijn pogingen om het menselijke gehoor te doorgronden en tijdens de talloze uren die ik doorbracht met het laten horen van biepjes aan proefpersonen, bleek het resultaat achteraf ook erg veel voldoening te geven. Later werd me ook duidelijk dat ExpORL een van de weinige labo's is wereldwijd waar zowel het technische als het medische aspect prominent aanwezig zijn, hetgeen het onderzoek zeer relevant en uniek maakt.

Veel dank ben ik verschuldigd aan mijn promotor, prof. Jan Wouters. Allereerst bood hij me de kans te starten met een doctoraat op een onderwerp dat me erg interesseerde, maar ook zorgde hij altijd voor optimale of zelfs ideale werkomstandigheden. Eenmaal de infrastructuur er is, bestaat het belangrijkste deel van een doctoraat echter uit mentale arbeid. Ook op dat gebied was Jan de ideale promotor. Hij had interesse in de kleinste nieuwe ontwikkeling of ontgoocheling, zijn kritische blik bracht tal van onvolkomenheden tijdig aan het licht en zijn ideeën hebben mijn onderzoek gestroomlijnd. Telkens ik weer eens iets te snel wou doen bracht Jan me terug op het rechte pad met zijn oog voor details. Niet alleen als "chef", maar ook als mens kon ik Jan zeer appreciëren, zijn gevoel voor humor en persoonlijke betrokkenheid zorgden voor een vrolijke noot, zowel in het lab als op talrijke uitstapjes met wetenschappelijke en minder wetenschappelijke doeleinden.

Prof. Marc Moonen wil ik bedanken voor het schijnen van zijn meer wiskundig licht op mijn onderzoek. In de gecompliceerde evenwichtsoefening tussen wiskundige signaalverwerking en klinische toepasbaarheid, wist hij me ten gepaster tijde in balans te houden. Ook mijn juryleden, prof. Haegemans, Van Compernolle, Van Hamme en in het bijzonder onze buitenlandse gasten dr. Brokx, prof. Kohlrausch en prof. Moore ben ik dankbaar voor hun kritische leeswerk en aanwezigheid bij de verdediging.

Hoorapparaten en cochleaire implantaten zijn enkel zinvol als ze gebruikt worden door mensen. Daarom is het bij onderzoek terzake onontbeerlijk om tests uit te voeren met mensen die deze toestellen gebruiken. Hoewel de medische wetenschap enorm vooruitgaat, blijven het menselijk lichaam en vooral de hersenen een onvoorspelbare factor, die soms frustrerend is, maar altijd fascinerend. Daarom ben ik mijn proefpersonen Annelies, Annemie, Bart, Chris, Frank, Gerard, Hanna, Jan, Kelly, Maria, Marinus, Myrthe, Pierre, Piet, René, Rob, Romain, Ruud, Sindy en Theo ook zeer dankbaar voor de vele uren die ze samen met mij doorbrachten, luisterend naar "biepjes" die van links of rechts kwamen of luider of stiller waren. Ondanks de saaie opdrachten waren ze altijd gemotiveerd en enthousiast. Niet alleen hebben de psychofysische experimenten waaraan ze deelnamen vruchten afgeworpen, maar ook heb ik van hen veel geleerd over hoe het leven is als slechthorende.

Bedankt Ann en Kathleen van het Revalidatiecentrum spraak en gehoor van het UZLeuven en Jan en Joke van het audiologisch centrum van het AZ Maastricht voor het leggen van de eerste contacten met proefpersonen. Bedankt ook Afra, Annemie, Ans, Audrey, Danielle, Els, Ester, Jacqueline, Jan, Joke, Lucien, Mirçea, Nadia, Peter, Sander, Sandra, Winde en Yvonne van het AZ Maastricht voor jullie belangeloze inzet, de vriendelijke opvang, flexibiliteit en fantastische samenwerking. Jan en Joke moet ik extra in de bloemetjes zetten voor hun blijvende inspanningen voor en betrokkenheid bij mijn onderzoek.

Cochlear en het IWT ben ik dankbaar voor de financiële steun. Cochlear zorgde ook voor technische ondersteuning in de vorm van een onderzoeksplatform en antwoorden op mijn vele vragen. In het bijzonder bedank ik Dieter Beaven, Wim Buyens, Colin Irwin, Bas Van Dijk en Clemens Zweekhorst voor de aangename en constructieve samenwerking.

In ons lab heerst altijd een erg aangename sfeer. Die zou er niet zijn zonder mijn toffe audiologie-collega's Bram, Hanne, Heleen, Jaime, Jane, Koen, Lot, Michael, Sofie en Tim en logopedie-collega's Catherine, Ellen, Eric, Evelyne, Inge, Joke, Stien, Tinne en Wivine. In het bijzonder wil ik Lot en Michael bedanken voor de fijne samenwerking en vriendschap en Koen voor de leuke en interessante muziek-discussies. Onmisbaar in

het lab is ook Frieda, die de administratieve molen punctueel en efficiënt draaiende houdt en iedereen een luisterend oor biedt. Ook Astrid was een belangrijke steun voor mij, door het supersnel nalezen van manuscripten, met waardevol wetenschappelijk advies en met haar zeer gesmaakte humor.

Het leven is meer dan werken alleen. Daarom gaat mijn dank ook uit naar mijn vrienden, ouders, grootouders en Annelies voor hun aanwezigheid en steun.

Abstract

A cochlear implant (CI) is a device that bypasses a nonfunctional inner ear and stimulates the auditory nerve with patterns of electric current, such that speech and other sounds can be experienced by profoundly deaf people. Due to the success of CIs, an increasing number of patients with residual hearing is implanted. In many cases they use a hearing aid (HA) in the non-implanted, severely hearing impaired ear. This setup is called bilateral bimodal stimulation. Despite the fact that binaural inputs are available, bimodal listeners exhibit poor sound source localization performance. This is partly due to technical problems with the processing in current CI speech processors and HAs.

Using an experimental setup, sensitivity was assessed to the basic localization cues, the interaural level difference (ILD) and interaural time difference (ITD). The just noticeable difference (JND) in ILD was measured in 10 bimodal listeners. The mean JND for pitch-matched electric and acoustic stimulation was 1.7 dB. However, due to insufficient high frequency residual hearing, users of bimodal aids do not have access to real-world ILD cues. Using noise band vocoder simulations with normal hearing subjects, it was shown that localization performance using bimodal aids can be improved by artificially amplifying ILDs in the low frequencies. Finally, the JND in ITD was assessed in 8 users of bimodal aids. Four subjects were sensitive to ITDs and exhibited JNDs in ITD of around $100 - 200\,\mu s$. The electric signal had to be delayed by on average 1.5 ms to achieve synchronous stimulation at the auditory nerves.

Overall, sensitivity to the binaural localization cues (ILD and ITD) was found to be well within the range of real-world cues. To allow the use of these cues for localization through clinical devices, they should be synchronized, matched in place of excitation and furthermore performance can be improved by ILD amplification in the low frequencies of the acoustic signal.

Korte Inhoud

Een cochleair implantaat (CI) is een apparaat dat het niet-functionele binnenoor omzeilt en de gehoorzenuw stimuleert met elektrische stroom zodat doven spraak en andere geluiden kunnen waarnemen. Door het succes van CIs, worden steeds meer patiënten geïmplanteerd die restgehoor hebben. In veel gevallen gebruikt deze groep een hoorapparaat (HA) in het niet-geïmplanteerde, ernstig slechthorende oor. Deze configuratie wordt bimodale stimulatie genoemd. Ondanks het feit dat er in dit geval binaurale informatie wordt aangeboden, scoren bimodale luisteraars slecht op geluidslokalisatietaken. Dit is deels te wijten aan technische problemen met de signaalverwerking in de CI spraakprocessor en het HA.

Met een experimentele opstelling werd de gevoeligheid nagegaan voor de basis cues voor lokalisatie van geluidsbronnen: het interauraal tijdsverschil (ITD) en het interauraal niveauverschil (ILD). Het kleinst waarneembare verschil (JND) in ILD werd gemeten bij 10 bimodale luisteraars. De gemiddelde JND voor in toonhoogte overeenkomende elektrische en akoestische stimulatie was 1.7 dB. Omwille van ontoereikend restgehoor bij hoge frequenties, hebben bimodale luisteraars echter geen toegang tot ILDs in realistische geluiden. Met behulp van ruisbandvocodersimulaties bij normaalhorende proefpersonen werd aangetoond dat het lokalisatievermogen met bimodale stimulatie verbeterd kan worden door versterking van ILDs bij lage frequenties. Tenslotte werd de JND in ITD opgemeten bij 8 bimodale luisteraars. Vier proefpersonen waren gevoelig voor ITDs met JNDs in ITD van $100 - 200\,\mu s$. Het elektrisch signaal moest gemiddeld met 1.5 ms worden vertraagd om synchrone stimulatie ter hoogte van de gehoorzenuwen te bereiken.

De gevoeligheid voor binaurale lokalisatie cues (ILD en ITD) was ruim binnen het bereik van realistische ILD en ITD cues. Om het gebruik van ILD en ITD mogelijk te maken met klinische toestellen, moeten deze gesynchroniseerd worden en overeengestemd in plaats van excitatie in de cochlea's. Het lokalisatievermogen kan verder verbeterd worden door ILDs te versterken in de lage frequenties van het akoestische signaal.

Glossary

ABR auditory brain stem response

AGC automatic gain control

AM amplitude modulated

AMT APEX Matlab Toolbox

BMLD binaural masking level difference

BTE behind the ear

CI cochlear implant

CVC consonant vowel consonant

EABR electrical auditory brain stem response

EAS electric acoustic stimulation

GUI graphical user interface

HA hearing aid

HRTF head related transfer function

ILD interaural level difference

ITD interaural time difference

ITE in the ear

JND just noticeable difference

LGF loudness growth function

NH normal hearing

MAA minimum audible angle

NIC nucleus implant communicator

PTA pure tone average

RMS root mean square

SNR signal to noise ratio

SRT speech reception threshold

XML extensible markup language

Contents

Voorwoord iii

Abstract vii

Korte Inhoud ix

Glossary xi

List of Figures xix

List of Tables xxv

Samenvatting 1
- 1 Motivatie . . . 1
- 2 Inleiding . . . 2
  - 2.1 Gehoorverlies . . . 2
  - 2.2 Hoorapparaten . . . 3
  - 2.3 Cochleaire implantaten . . . 4
  - 2.4 Bimodale stimulatie . . . 6
  - 2.5 Lokalisatie van geluidsbronnen . . . 7
- 3 Het testplatform: APEX 3 . . . 9
- 4 Perceptie van ILDs over frequentiegrenzen heen . . . 9
- 5 Perceptie van ILDs met bimodale stimulatie . . . 10
- 6 Versterking van ILDs in de lage frequenties . . . 10
- 7 Perceptie van ITDs met bimodale stimulatie . . . 11
- 8 Besluit . . . 11

1 Introduction 13
- 1.1 Motivation . . . 13
- 1.2 Hearing loss . . . 14
- 1.3 Hearing aids . . . 15
- 1.4 Cochlear implants . . . 17
  - 1.4.1 The speech processor . . . 19
  - 1.4.2 Design issues . . . 20

1.4.3 Subject performance with CIs . . . 22
1.5 Bimodal stimulation . . . 23
1.5.1 Problems with current clinical bimodal systems . . . 24
1.5.2 Matching the place of excitation . . . 26
1.6 Localization . . . 27
1.6.1 Measuring localization performance . . . 28
1.6.2 Localization error . . . 31
1.6.3 Minimum audible angle . . . 31
1.6.4 ILD . . . 32
1.6.5 ITD . . . 33
1.6.6 Monaural and visual cues . . . 34
1.6.7 Head related transfer functions . . . 35
1.6.8 Adaptation to changes in localization cues . . . 35
1.7 Thesis outline . . . 36

2 APEX 3: a test platform for auditory psychophysical experiments 39
2.1 Introduction . . . 39
2.1.1 The hardware platform . . . 40
2.1.2 The software platform: APEX 3 . . . 41
2.2 Concepts . . . 44
2.3 Design . . . 46
2.3.1 ApexControl . . . 47
2.3.2 Procedure . . . 47
2.3.3 Device . . . 49
2.3.4 Controller . . . 49
2.3.5 Screen . . . 50
2.3.6 ResultSink . . . 50
2.3.7 Calibrator . . . 51
2.3.8 Filters . . . 51
2.3.9 Connections . . . 51
2.4 Extending APEX 3 . . . 52
2.4.1 PluginProcedure . . . 52
2.4.2 PluginController . . . 53
2.4.3 PluginFilter . . . 53
2.5 Defining an experiment . . . 53
2.5.1 A simple example experiment . . . 55
2.5.2 Writing experiment files . . . 59
2.6 Workflow . . . 60
2.7 Examples . . . 61
2.7.1 Gap detection using a 3-AFC paradigm with a CI . 61

2.7.2 Adaptive determination of the SRT 63
2.7.3 Automatic determination of the SRT 65
2.7.4 Evaluation of a signal processing algorithm with an adaptive SRT procedure 65
2.7.5 Bimodal stimulation 65
2.7.6 Localization of sounds 67
2.8 Conclusions 67

3 Across-frequency ILD perception in normal hearing subjects 69
3.1 Introduction 70
3.2 Methods 71
3.2.1 Procedure 71
3.2.2 Experimental setup 73
3.3 Results 75
3.3.1 Experiment 1 75
3.3.2 Experiment 2 75
3.4 Discussion 78
3.5 Conclusions 80

4 ILD perception and loudness growth with bimodal stimulation 83
4.1 Introduction 83
4.2 Methods 85
4.2.1 Apparatus 85
4.2.2 Stimuli 85
4.2.3 Procedures 86
4.2.4 Subjects 92
4.3 Results 93
4.3.1 Pitch matching 93
4.3.2 Loudness growth functions and JNDs in ILD 95
4.4 Discussion 96
4.4.1 Pitch matching 96
4.4.2 Loudness growth 101
4.4.3 Just noticeable differences 102
4.4.4 Relation to localization performance 102
4.5 Conclusions 105

5 ILD amplification for bilateral bimodal stimulation 107
5.1 Introduction 108
5.2 General Methods 108
5.2.1 Simulation of directional hearing 108
5.2.2 Signals 109

5.2.3 Apparatus . . . 111
5.2.4 Subjects . . . 111
5.3 Experiment 1 . . . 112
5.3.1 Methods . . . 112
5.3.2 Results . . . 116
5.3.3 Discussion . . . 118
5.4 Experiment 2 . . . 122
5.4.1 Methods . . . 122
5.4.2 Results . . . 129
5.4.3 Discussion . . . 129
5.5 General discussion and conclusions . . . 131

6 ITD perception with bilateral bimodal stimulation 133
6.1 Introduction . . . 134
6.2 Methods . . . 135
6.2.1 Apparatus . . . 135
6.2.2 Stimuli . . . 135
6.2.3 Procedures . . . 138
6.2.4 Subjects . . . 144
6.3 Results . . . 145
6.3.1 Fusion . . . 145
6.3.2 Loudness balancing and intensity . . . 146
6.3.3 JND in ITD . . . 147
6.3.4 Delays . . . 153
6.3.5 Matching the place of excitation . . . 153
6.4 Discussion . . . 154
6.4.1 Fusion . . . 154
6.4.2 Influence of ILD . . . 154
6.4.3 JND in ITD . . . 154
6.4.4 Delays . . . 155
6.4.5 Relation with localization performance and binaural unmasking . . . 156
6.5 Conclusions . . . 157

7 Conclusions and further research 159
7.1 Conclusions . . . 159
7.1.1 ILD sensitivity . . . 159
7.1.2 Improving localization by ILD amplification . . . 160
7.1.3 ITD sensitivity . . . 160
7.1.4 Impact on localization performance and binaural unmasking . . . 161

7.2 Further research 161
7.2.1 Further psychophysical research 161
7.2.2 Further technical research 162

A Automatic testing of speech recognition **165**
A.1 Introduction 166
A.2 Description of the algorithms 167
A.2.1 The sentence algorithm 169
A.2.2 The word algorithm 176
A.3 Evaluation of the algorithms 179
A.3.1 Development of a test corpus: procedures 180
A.3.2 Materials 181
A.3.3 Subjects 182
A.3.4 Evaluation 182
A.4 Results 184
A.5 Discussion 187
A.6 Conclusion and applications 189

B Roving for across frequency ILD perception **191**

C Roving for ILD amplification experiments **193**
C.1 Introduction 193
C.2 Methods 195
C.3 Results and discussion 195
C.4 Conclusions 197

Bibliography **198**

List of publications **213**

Curriculum Vitae **217**

List of Figures

1.1 Median hearing thresholds versus age for otologically normal male subjects (from ISO-7029) 14
1.2 I/O characteristics of an example automatic gain control (AGC). In this figure a wide dynamic range compression characteristic is shown. The first part is linear, followed by a non-linear part from input level 40 dB SPL with compression ratio 3, and output clipping at levels higher than 100 dB SPL. 16
1.3 General overview of the internal and external parts of a cochlear implant system (reprinted with permission from Cochlear) 18
1.4 Drawing of an electrode array implanted in a cochlea (reprinted with permission from Cochlear) 18
1.5 Block diagram of the signal processing in the ACE strategy 19
1.6 CI compression characteristics for different values of Q . . . 21
1.7 Illustration of synchronization problems in bimodal systems 24
1.8 Illustration of place mismatch between electric and acoustic stimulation 24
1.9 Illustration of ILD and ITD for a sound incident from the left side of the head (-90°) 27
1.10 Schematic of the localization test setup at ExpORL 28

2.1 Experimental setup for synchronized electric acoustic stimulation 40
2.2 Overview of the general working of Procedure. Procedure presents a trial by selecting a Stimulus to be sent to the stimulus output logic and a Screen to be shown. 46
2.3 Overview of several APEX 3 modules. The stimulation box is not an APEX 3 module, but groups all stimulation-related modules. The four bottom right boxes do not show a complete description of datablocks, stimuli, devices and screens, but serve to guide the eye and indicate that the corresponding modules are defined. 47

2.4 Connection graph of the simple example, as generated by APEX 3. In this case each datablock has two channels (left and right) that are connected to the two channels of the sound card. The left and right channels are indicated by the numbers 0 and 1, respectively. 52
2.5 Screen of the example experiment 58
2.6 Workflow conducting an experiment using APEX 3. AMT is the APEX 3 Matlab Toolbox. 60
2.7 Example of an arcLayout with $N = 9$ buttons 67

3.1 Example of a standard-stimulus sequence with a positive rove. For this trial, the correct answer would be "The stimulus sounded on the left hand side of the standard". 72
3.2 All normalized sequences of runs for experiment 1. All values for each sequence were divided by the mean of the last 6 runs for that sequence. Each dot represents the result of an adaptive run. The full line connects the averages at each time instant. 76
3.3 JNDs in ILD (in dB) as a function of base frequency and frequency shift for experiment 1 (±5 dB rove). The total length of the error bar is twice the standard deviation. The data were checked for normality using the Kolmogorov-Smirnov test. 77
3.4 Differences between experiment 1 and 2. The bars show the difference in JND. The error bars represent the combined error of both experiments. Positive values indicate that the JND in experiment 1 (±5 dB rove) was larger than the JND in experiment 2 (±10 dB rove). 79

4.1 Psychometric function for the fine pitch matching experiment for subject S4, set 1. 89
4.2 Example psychometric function for a loudness balancing experiment for S2, set 2 with the acoustic level fixed at 110 dB SPL. The JND in ILD was 6.5 % of the electric dynamic range and 64 % of the electric dynamic range corresponded to an acoustical intensity of 110 dB SPL. 91
4.3 Unaided subject audiograms for the acoustically stimulated ear. Note that the vertical axis starts at 50 dBHL. No symbol means no threshold could be found at that frequency. 94

4.4 Matched pitches for the most apical electrode per subject. Note that S9 has a partial electrode insertion, which explains the higher pitch. 95

4.5 loudness growth functions (LGFs) between electrical and acoustical stimulation for set 1. The error bars were determined using the bootstrap method. 97

4.6 LGFs between electrical and acoustical stimulation for set 2. The error bars were determined using the bootstrap method. 98

4.7 JNDs for each electrical intensity for subject S6. The X-axis shows the fraction of the electrical dynamic range in current units. Error bars are 68% confidence intervals determined using a bootstrap method. The dashed line shows the median and the thick error bar on the right hand side shows the 25% and 75% quartiles. 99

4.8 All JNDs in ILD expressed as dB change in the acoustical ear for a fixed electrical current in the other ear. The 75% and 25% quantiles are indicated. The JND for S10, set 2 is 7.6 dB. Above the label 2CI, the diamonds show data from Senn et al. (2005) and the plusses show data from Laback et al. (2004) for bilateral CI users. Above the label NH, the diamonds show data from Yost and Dye (1988) and the plusses show data from Mills (1960) for normal hearing listeners. 100

4.9 Simulated transfer function of the SPrint processor for two different fittings. The abrupt changes in the function are due to quantization into current levels and the breakpoint is due to the saturation level implemented in the loudness growth processing. 103

5.1 Spectrum of the telephone signal 110

5.2 ITDs per gammatone filter of the phase randomization filter, determined from the maximum of the cross correlation function between the left and right channel. Every symbol shows the ITD between the corresponding channels of the gammatone filter bank in the left and right ear. The correlation between the two channels was much lower than before application of the phase randomization filter and the cross correlation method did not yield meaningful ITDs. . . 114

5.3 Overview results of experiment 1 part 1 – P_{FB} for each of 11 normal hearing subjects. The error bars represent standard deviations on the average of test and retest. 117
5.4 Average results for experiment 1 over all subjects (test and retest). The error bars show standard deviations. RMS errors lower than 67.3° are significantly better than chance level (indicated by the dashed line). 118
5.5 Results for Experiment 1 for each angle, averaged over all subjects (test and retest). The error bars are between-subject standard deviations. 119
5.6 ILD for each frequency and angle of incidence, determined from ITE HRTFs, measured using an artificial head. 120
5.7 Analysis and synthesis filters used for the noise band vocoder CI simulation 124
5.8 Sixth order Butterworth filter used to simulate severe hearing loss 125
5.9 Levels of the wide band signals (noise14000 and telephone) after filtering with BTE head related transfer functions (HRTFs), with and without simulation of bimodal hearing, before and after application of the ILD amplification algorithm. The noise band vocoder simulation (CI) was done for the left ear and the low pass filtering (HA) for the right ear. The ILD at a certain frequency can be obtained by subtracting the respective levels in dB for the left and right ears. 127
5.10 The top panel shows the results of experiment 2, averaged for each signal and condition. amp is the condition with application of the ILD amplification algorithm and noamp is the condition without ILD amplification. The bottom panel shows the same results per angle of incidence. 130

6.1 Part of an example stimulus. The top panel shows a filtered click train with harmonics 2-4 and $F_0 = 100$ Hz and the bottom panel shows an electric pulse train of 100 pps. . . . 136
6.2 Part of an example transposed stimulus. The top panel shows a transposed sinusoid with a base frequency of 1000 Hz and a modulation frequency of 42 Hz. The bottom panel shows an electric pulse train of 6300 pps modulated with a half wave rectified sinusoid with a frequency of 42 Hz. . . . 137

6.3 Graphical overview of the used fusion and loudness balancing procedures. The white boxes illustrate the stimuli presented to the subject and the text on the right shows example results from each procedure. Each procedure used parameters determined in the previous procedure, which are shown with a gray background. The numbers are fictive but of a realistic magnitude. The plotted electric and acoustic signals only serve illustrative purposes and only show parts of example signals. 140

6.4 Example psychometric function for S4 used to determine the JND in ITD and delay required for psychoacoustically synchronous stimulation. using electrode 11 and harmonics 8-16 for the acoustic signal. The level of the acoustic signal was 100 dB SPL and the level of the electric signal was 45 % of the dynamic range. From the found crossover point ($-3112\,\mu s$), the delay of the used insert phone (1154 µs) has to be subtracted to find D_e. For the measurement of this psychometric function, 63 trials were used. 143

6.5 Unaided pure tone audiograms for each subject as measured during routine audiometry. Note that the vertical axis starts at 60 dB HL. If no symbol is shown, no threshold could be measured using the clinical audiometry equipment. 144

6.6 JND in ITD in µs for each subject and condition. A cross indicates that the condition was tested, but that sensitivity to ITD was insufficient to do the lateralization task. 149

6.7 Best median JND in ITD per subject and per electrode. The values above the label "2x CI" are reference values from the bilateral CI literature for pulse trains of 100 pps. Each symbol is the JND in ITD for one subject. The error bars are 68% confidence intervals on the fit of the psychometric function, determined by a bootstrap method. 150

6.8 ITD perception performance versus thresholds of residual hearing. Each different symbol denotes a different threshold measurement frequency. The filled circles show the average threshold at frequencies 1000 and 2000 Hz. 151

6.9 Histogram of D_e values. Each value contributing to an increment of one on the vertical axis corresponds to a value found by fitting a psychometric function to the response for between 21 and 117 trials. If measurements with the same stimulus were available at different ILDs, only the D_e was selected for which the corresponding JND in ITD was smallest. 152

A.1 Flowchart of the sentence algorithm. An arrow signifies that the output from the source block is used as the input for the target block. 170
A.2 Example of string alignment. Spaces are marked by empty boxes. In this case the gold string is "word score" and the user input string "woldsc re". First all spaces are removed from the input string. Then both strings are aligned. The space character marked with the single arrow could be inserted into the input string as shown. However, as the percentage of correctly aligned characters ($100 \cdot 7/10 = 70\%$) is smaller than 90%, no space will be inserted because the strings are not considered sufficiently alike in this case. . . 174
A.3 General structure of the word correction algorithm. 177

C.1 Monte Carlo simulations of average RMS error obtained with a decision strategy only using monaural loudness cues for different roving ranges (R). Each data point is the median of 10^5 simulations. 194
C.2 Comparison of binaural and monaural results. The dotted and dashed lines show the significance and chance level, respectively. 196

List of Tables

4.1 Subject information: "Age" is the age in years at the time of testing. "M of use" is the number of months of implant use at the time of testing. "CI side" is left (L) or right (R) (the HA was on the other side). "Elec" is the electrode number (numbered from apex to base) and "DR" is the electrical dynamic range in current units. "MF" is the frequency of the pitch matched sinusoid in Hz and "Thr" is the acoustical threshold in dBSPL 87

5.1 Overview of the signals used. 111

6.1 Subject information: Age is in years at the time of testing. M of use is the number of months of implant use at the time of testing. CI side is left (L) or right (R). The HA was on the other side. Perf is the category of ITD perception performance. A, M and B are the tested electrodes at apical, medial and basal positions in the electrode array. 145
6.2 JNDs in jitter balancing and percentages of jitter for the acoustic (Ac) and electric (El) signal used for the subsequent balancing experiment. 146
6.3 Range of harmonics of best matching acoustic signal for each electrode and subject. If there was no clear difference between two acoustic signals, both are given. 148
6.4 Wave V latencies (in ms) from different studies on ABR and EABR. All were measured at a comfortably loud level. The last row shows reference values used for the clinical ABR setup in our hospital (UZLeuven). 156

A.1 Scoring rules for CVC tests and sentence tests 168
A.2 Description of regular expressions used for the Dutch LIST and VU sentence test materials 171
A.3 Example values of $H(w)$ for single and double characters, a long word, and a bigram. [From equation A.1 with $N = 5$] 173

A.4 Bigrams and corresponding hash values for "the boy fell from the window" . . . 176
A.5 Bigrams and corresponding hash values for bigrams can be corrected from the user input sentence. (a bigram can only be corrected if it contains a word that is not found in the dictionary) . . . 177
A.6 Graphemes used for correction of Dutch CVC words. Graphemes with the same code are between square brackets and codes are given as subscripts. . . . 179
A.7 Percentage of errors made by the autocorrection algorithm compared to manual scoring methods for each speech material, group of subjects (Grp), number of tokens in the corpus and corpus entry type. For the sentence materials, errors for keyword score (Word) and for sentence score (Sent) are given. For the CVC materials, errors for phoneme score and for word score are given. # is the total number of sentences presented for the sentence tests and the total number of words presented for the CVC test. MO-MOC is the percentage of changes between the MO and MOC scores. ΔSRT is the mean of the differences in estimated SRT (in dB) between Ac and MOC for each condition. MO is the original manual score based on the oral response, MOC is the corrected manual score based on the oral response, MT is the manual score based on the typed response and Ac is the score by the autocorrection algorithm. . . . 183

C.1 Localization cues available in the different conditions. . . . 195

Perceptie van binaurale lokalisatiecues met gecombineerd elektrisch en akoestisch gehoor

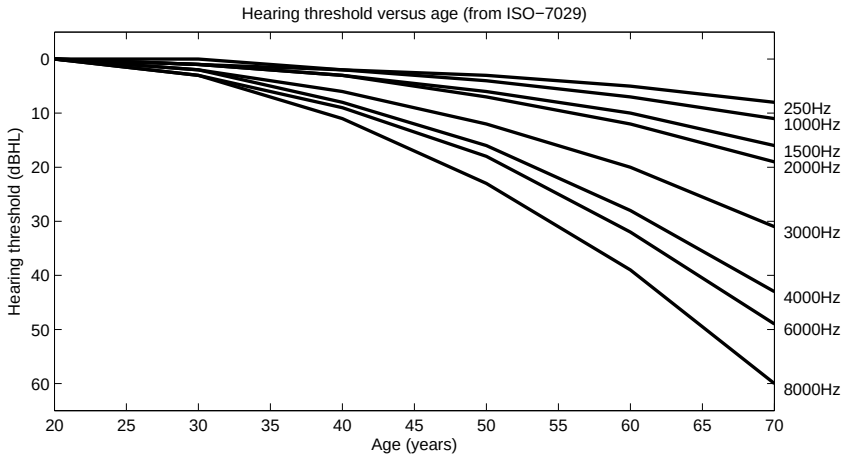
Uitgebreide samenvatting in het Nederlands

1 Motivatie

Een cochleair implantaat (CI) is een apparaat dat het disfunctionele gehoor omzeilt en de gehoorzenuw stimuleert met elektrische stroom. Zo kunnen volledig doven terug spraak verstaan of andere geluiden waarnemen. Dankzij de vooruitgang van CIs, kan bij ernstig slechthorenden een CI tot beter spraakverstaan leiden dan een hoorapparaat (HA). Daarom zijn er steeds meer patiënten met restgehoor in een van beide oren die een CI krijgen. We zullen ons toespitsen op het geval van CI-gebruikers met restgehoor in het niet-geïmplanteerde oor. De situatie waarbij iemand een CI gebruikt samen met akoestisch gehoor noemt men *bimodale* stimulatie of elektrisch akoestische stimulatie. Gezien de hoge kostprijs van een tweede implantaat[1] en eventuele voordelen verbonden aan bimodale stimulatie, hebben we te maken met een groeiende populatie patiënten die een CI gebruiken in het ene oor en een hoorapparaat (HA) in het andere.

Normaalhorenden (NH) lokaliseren geluidsbronnen vooral door het vergelijken van de geluiden tussen beide oren. Aangezien bij bimodale stimulatie beide oren gestimuleerd worden, zou men kunnen verwachten dat het lokaliseren van geluidsbronnen veel beter gaat dan met enkel een CI in één oor. Er werd inderdaad aangetoond dat bij veel proefpersonen de lokalisatie verbetert bij de combinatie van een HA met het CI (Ching et al., 2007). Toch scoren ze in vergelijking met NH proefpersonen nog zeer slecht op lokalisatietaken. Dit is te wijten aan verschillende technische problemen met de CIs en HAs die op dit moment klinisch gebruikt worden.

[1] Standaard wordt in de meeste landen slechts één CI terugbetaald door de ziekteverzekering. Gezien de hoge kostprijs van een CI hebben de meeste CI gebruikers dan ook slechts een CI in één van beide oren.



Figuur 1: Gemiddelde gehoordrempel versus leeftijd voor otologisch normale mannen (uit ISO-7029)

Het doel van dit doctoraat was om na te gaan of gebruikers van een bimodaal systeem de basisinformatie, nodig voor het lokaliseren van geluidsbronnen, kunnen waarnemen indien de technische problemen opgelost of omzeild worden.

2 Inleiding

2.1 Gehoorverlies

Gehoorverlies is een veel voorkomende handicap. Volgens de World Health Organization (WHO) hadden in 2005 wereldwijd 278 miljoen mensen een matig tot zeer ernstig gehoorverlies. In de ISO-7029 standaard worden gemiddelde gehoordrempels gegeven voor de otologisch normale populatie. In figuur 1 wordt als voorbeeld de mediaan van de gehoordrempel[2] getoond voor de mannelijke bevolking per leeftijdscategorie. Het is duidelijk dat een groot deel van de bevolking gedurende het leven met gehoorverlies

[2]De gehoordrempel is het niveau van het zachtste geluid dat wordt waargenomen, opgemeten per frequentie. Klassiek wordt de gehoordrempel opgemeten op 125, 250, 500, 1000, 2000, 4000 en 8000 Hz en uitgedrukt in decibel (dB HL). Normaalhorenden hebben op alle frequenties een gehoordrempel van $0 - 20$ dB HL. Bij slechthorenden kan dit oplopen tot 120 dB HL of zelfs onmeetbaar zijn.

Type	PTA
Licht	$25 - 40\,\text{dB HL}$
Matig	$40 - 55\,\text{dB HL}$
Matig-ernstig	$55 - 70\,\text{dB HL}$
Ernstig	$70 - 90\,\text{dB HL}$
Zeer ernstig	$> 90\,\text{dB HL}$

Tabel 1: Categorieën van gehoorverlies

geconfronteerd wordt.

Er zijn vele mogelijke oorzaken van doofheid. Het kan erfelijk zijn, maar ook verworven, bijvoorbeeld door langdurige blootstelling aan lawaai, veroorzaakt door machines of door langdurige blootstelling aan luide muziek bijvoorbeeld door een MP3 speler (Fligor and Cox, 2004; LePage and Murray, 1998; Williams, 2005).

Gehoorverlies wordt in verschillende categorieën opgedeeld aan de hand van de pure tone average (PTA)[3]. De verschillende categorieën worden gegeven in tabel 1. We zullen ons vooral toespitsen op mensen met ernstig tot zeer ernstig gehoorverlies.

2.2 Hoorapparaten

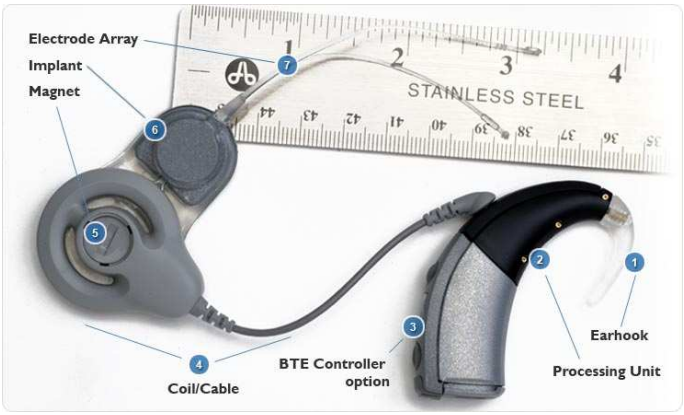
Een belangrijk probleem bij gehoorgestoorden is gereduceerde hoorbaarheid. Zachte geluiden worden niet meer waargenomen. Dit kan in een hoorapparaat (HA) eenvoudig opgelost worden door alle binnenkomende geluiden te versterken. Een HA in zijn meest eenvoudige vorm is dan ook een versterker[4]. Dit levert echter een nieuw probleem op: harde geluiden worden onaangenaam hard. De drempel voor onaangenaam luide geluiden schuift niet samen op met de gehoordrempel. Het gevolg is dat het dynamisch bereik[5] gereduceerd wordt. Bij ernstig tot zeer ernstig gehoorverlies kan het dynamisch bereik beperkt zijn tot slechts 10 dB, terwijl NH een dynamisch bereik hebben van meer dan 100 dB.

Om dit probleem op te lossen, reduceren HAs het dynamisch bereik

[3] De PTA is de gemiddelde gehoordrempel voor zuivere tonen op 500, 1000 en 2000 Hz

[4] De belangrijkste parameter van een versterker is de *versterking*, meestal uitgedrukt in decibel (dB), die aangeeft in welke mate het signaal versterkt wordt.

[5] Het dynamisch bereik is het geheel van geluidsintensiteiten dat kan worden waargenomen. Het komt overeen met het verschil tussen de gehoordrempel en de drempel voor onaangenaam luide geluiden en wordt bijgevolg uitgedrukt in decibel.



Figuur 2: Overzicht van het interne en externe deel van een CI (overgenomen met toestemming van Cochlear)

met behulp van AGC[6]. AGC past de versterking van het HA aan aan het *gemiddelde* geluidsniveau gedurende zekere tijd. Zo zullen zachte geluiden meer versterkt worden dan harde geluiden.

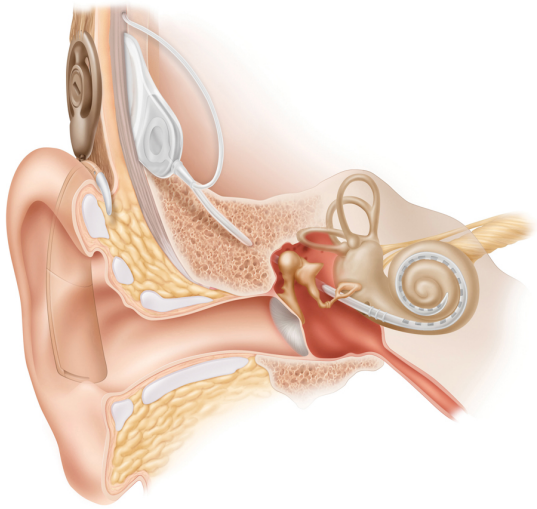
Aangezien HAs het functionele uitwendige en inwendige oor benutten, zijn ze niet nuttig voor volledig dove patiënten.

2.3 Cochleaire implantaten

Een cochleair implantaat (CI) is een apparaat dat het disfunctionele gehoor omzeilt en de gehoorzenuw stimuleert met elektrische stroom. Er wordt geschat dat wereldwijd 100.000 mensen gebruik maken van een CI. Volgens data verzameld door EuroCIU waren er in 2007 in Europa meer dan 47.000 CI-gebruikers, waarvan 1000 bilateraal[7] geïmplanteerd en waarvan meer dan 22.000 kinderen. In België waren er 1620 CI-gebruikers. Op de CI-markt zijn drie grote bedrijven actief: Cochlear, Advanced Bionics en Med-El. Met zijn 65-70% marktaandeel is Cochlear de grootste speler wereldwijd.

Een CI bestaat uit twee belangrijke delen: het externe en het interne deel (zie figuur 2 en 3). Het externe deel wordt achter het oor geplaatst, zoals een HA. Via een antenne (een spoeltje) wordt informatie doorgestuurd naar het interne deel. Het interne deel wordt tijdens een operatie

[6] automatische instelling van de versterking
[7] aan beide oren



Figuur 3: Tekening van een elektrode-array geïmplanteerd in de cochlea (overgenomen met toestemming van Cochlear)

geïmplanteerd. Het bestaat uit een ontvanger/stimulator en een electrode-array[8]. De electrode-array wordt ingebracht in de cochlea[9] tijdens een operatie (zie figuur 3).

Het externe deel wordt ook de spraakprocessor genoemd. Het bevat elektronica die de geluiden, die invallen op de microfoon, omzet in series van elektrische pulsen die in het interne deel naar de verschillende elektroden gestuurd worden. Er zijn vele manieren om deze omzetting te doen. Zo'n manier noemt men een spraakverwerkingsstrategie. Een voorbeeld van een veel gebruikte spraakverwerkingsstrategie is de ACE[10] strategie, zoals geïmplementeerd in de spraakprocessoren van Cochlear.

Een belangrijke eigenschap van de spraakverwerkingsstrategie is de manier waarop een bepaalde akoestische frequentie toegekend wordt aan één van de elektrodes. De cochlea heeft een tonotopische organisatie, dit betekent dat elke plaats in de cochlea overeenkomt met een bepaalde frequentie

[8] De electrode-array, ook elektrode-strip genoemd, is een serie van elektrodes (12 à 22, naargelang het type en de producent) die op gelijke afstanden van elkaar op een drager zijn aangebracht.

[9] De cochlea, ook slakkenhuis genoemd, maakt deel uit van het binnenoor en is de interface tussen geluidsgolven en stimuli op zenuwvezels.

[10] Advanced Combination Encoder

van een akoestisch signaal. Indien de plaats van de elektrodes die door de spraakverwerkingsstrategie gekozen worden ver afwijkt van de plaats waar een geluid "normaal" zou worden aangeboden, is een lang leerproces (tot zelfs een jaar lang) nodig om tot maximaal spraakverstaan te komen.

Een spraakverwerkingsstrategie heeft een aantal instellingen die per patiënt worden ingesteld om maximaal comfort en spraakverstaan te bereiken. Dit wordt gedaan door een audioloog tijdens de zogenaamde "fitting", meestal in een revalidatiecentrum verbonden aan het ziekenhuis waar de implantatie uitgevoerd werd.

Aangezien CI-gebruikers net zoals HA-gebruikers slechts een beperkt dynamisch bereik hebben (zie sectie 2.2), bevat ook een spraakprocessor AGC.

CIs zijn zeer succesvol in die zin dat ze volledig doven toelaten om terug spraak te verstaan. De mate van succes is echter erg afhankelijk van de individuele patiënt. Dit is deels te verklaren door verschillende etiologieën en deels door andere factoren, zoals bijvoorbeeld cognitieve. De meeste patiënten zijn echter goed in het verstaan van spraak in stilte. In achtergrondlawaai is er veel meer variatie, maar alle patiënten doen het een stuk slechter dan normaalhorenden.

2.4 Bimodale stimulatie

Indien iemand gebruik maakt van een CI in het ene oor en een HA in het andere, spreekt men van bimodale stimulatie. Deze configuratie komt steeds meer voor omdat, gezien het succes van CIs, de implantatiecriteria verzwakken en omdat een tweede CI erg duur is. Er is nu dus sprake van een groeiende populatie patiënten die een CI gebruiken in het ene oor en restgehoor hebben in het andere oor, meestal enkel op de lage frequenties ($< 1000\,\text{Hz}$). Dit restgehoor wordt benut door middel van een HA.

In de huidige klinische praktijk krijgen mensen een CI en HA die enerzijds vrijwel volkomen los van elkaar ontwikkeld zijn en anderzijds soms ook onafhankelijk worden ingesteld (gefit). Aangezien deze toestellen niet ontwikkeld zijn om samen te werken, geeft dit aanleiding tot een aantal technische problemen. De voornaamste problemen zijn binaurale luidheidsaangroei, synchronisatie, "place mismatch" en geluidskwaliteit. Deze problemen worden besproken in de volgende paragrafen.

binaurale luidheidsaangroei In de spraakprocessor wordt de luidheid van een geluid "vertaald" naar een elektrische stroomsterkte. Dit proces is geoptimaliseerd voor spraakverstaan, maar heeft niet per se hetzelfde effect als "normaal" gehoor. De spraakprocessor bevat bo-

vendien AGC die onafhankelijk werkt van de AGC in het HA. Beide processen samen hebben als gevolg dat een verschil in geluidsniveau vertaald wordt in een bepaald perceptueel verschil in het elektrisch gestimuleerde oor en een ander perceptueel verschil in het akoestisch gestimuleerde oor. Dit heeft nadelige effecten op de perceptie van interaurale niveauverschillen (zie sectie 2.5).

synchronisatie De spraakprocessor heeft een zekere tijd (vertraging) nodig om een geluid om te zetten in elektrische stroom. Ook het HA heeft een zekere tijd (vertraging) nodig om een geluid te versterken. Deze vertragingen komen in de meeste gevallen niet overeen. Bovendien heeft het geluid aan de HA kant tijd nodig om via het midden- en binnenoor de gehoorzenuw te bereiken, terwijl dit bij elektrische stimulatie onmiddellijk gebeurt. Deze verschillen in vertraging tussen het akoestische en het elektrische pad hebben nadelige effecten op de perceptie van interaurale tijdsverschillen (zie sectie 2.5).

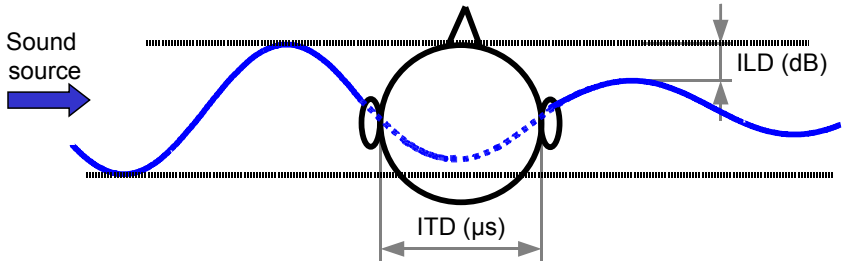
place mismatch Naargelang de frequentie-inhoud wordt een akoestisch signaal op een bepaalde plaats in de cochlea opgevangen. De plaats waar een elektrisch signaal terechtkomt, wordt echter bepaald door de spraakprocessor en zal in huidige systemen in veel gevallen niet overeenkomen met de akoestische plaats. Dit heeft nadelige effecten voor onder andere perceptie van interaurale niveau- en tijdsverschillen.

geluidskwaliteit De geluidskwaliteit van een CI is totaal anders dan die van akoestische stimulatie. Dit is oncomfortabel voor de gebruiker en heeft een negatieve invloed op de binaurale integratie van geluiden.

2.5 Lokalisatie van geluidsbronnen

Om de plaats van een geluidsbron te bepalen maakt het binauraal systeem gebruik van twee belangrijke aanwijzingen (zogenaamde *cues*). De eerste cue is het interaurale niveauverschil (ILD), het verschil in geluidssterkte tussen de oren. Als een geluid van links invalt, valt het rechtstreeks in op het linkeroor, en zal het door het *hoofdschaduweffect* in niveau verzwakt worden voor het het rechteroor bereikt (zie figuur 4). Dit niveauverschil is richtingafhankelijk en kan gebruikt worden om de plaats van de bron te bepalen. De tweede cue is het interaurale tijdsverschil (ITD). Als een geluid van links invalt, bereikt het eerst het linkeroor en pas enige tijd[11]

[11] Naargelang de richting van het invallende geluid, varieert de ITD van 0 tot ongeveer 700 µs



Figuur 4: Grafische voorstelling van ILD en ITD voor een geluidsbron aan de linkerkant van het hoofd

later het rechteroor (zie figuur 4). Dit tijdsverschil is richtingafhankelijk en kan gebruikt worden om de plaats van de bron te bepalen.

Omwille van fysische eigenschappen van het geluid en het hoofd en de oorschelpen, zijn ILDs fysisch vooral aanwezig op hogere frequenties (> 1500 Hz) en zijn ITDs enkel bruikbaar op lage frequenties (< 1500 Hz). Het menselijk gehoor is echter gevoelig voor ILDs over het hele frequentiebereik. Het gehoor is gevoelig voor zowel ITDs op lage frequenties als voor ITDs in de omhullende van een hoogfrequent signaal.

Het lokalisatievermogen kan op verschillende manieren gemeten worden. Eén mogelijkheid is om een proefpersoon in een kamer met verschillende luidsprekers te zetten en te vragen uit welke luidspreker het geluid komt. Anderzijds is het mogelijk om de gevoeligheid voor de binaurale cues (ILD en ITD) te meten. Dit kan men doen door via een hoofdtelefoon geluiden aante bieden en te vragen aan welke kant ze "gelateraliseerd" worden. Aangezien signalen gespeeld door een hoofdtelefoon gewoonlijk binnen het hoofd waargenomen worden, i.e., niet geëxternaliseerd worden, spreekt men hier over lateralisatie in plaats van lokalisatie.

De gevoeligheid voor de binaurale cues wordt typisch uitgedrukt als een just noticeable difference[12] (JND). Men kan de JND definiëren als het verschil dat nog in 50% van de gevallen wordt waargenomen. Zo is bijvoorbeeld voor NH de JND in ILD voor zuivere tonen ongeveer 1 dB. Dit betekent dat als 100 keer één van 2 geluiden wordt aangeboden, waarin al dan niet een ILD van 1 dB zit, en de proefpersoon moet antwoorden of er al dan niet een verschil is, de proefpersoon ongeveer 50 keer van de 100 juist zal antwoorden.

In dit doctoraat werd de gevoeligheid voor binaurale cues van bimodale

[12]kleinst merkbare verschil

luisteraars opgemeten en werden verbeteringen voorgesteld aan de CI- en HA-signaalverwerking zodat de binaurale cues gebruikt kunnen worden voor lokalisatie van geluidsbronnen in reële situaties.

3 Het testplatform: APEX 3

Om de gevoeligheid voor binaurale cues op te meten is er volledige controle nodig over zowel de akoestische als de elektrische stimulus. Daarom gebruikten we geen klinische CI-spraakprocessoren en HAs, maar een insert phone voor de akoestische stimulatie en een experimentele spraakprocessor met rechtstreekse verbinding met de computer.

Om psychofysische experimenten uit te voeren, zoals bijvoorbeeld het bepalen van een JND in ILD of ITD, ontwikkelden we een experimenteel softwareplatform genaamd APEX 3. Dit platform ondersteunt zowel akoestische, bimodale als bilaterale CI-stimulatie en om het even welke psychofysische procedure kan geïmplementeerd worden. De ontwikkeling en werking van dit platform wordt beschreven in hoofdstuk 2.

Dit onderzoeksplatform wordt gratis beschikbaar gesteld aan onderzoeksinstellingen. Het wordt ondertussen voor verschillende studies gebruikt bij ExpORL en tevens in verschillende internationale laboratoria.

4 Perceptie van ILDs over frequentiegrenzen heen

Om de invloed na te gaan van place mismatch (zie sectie 2.4) op de perceptie van ILDs, voerden we een experiment uit met 12 NH. We bepaalden de JND in ILD voor stimuli waarbij het geluid in het ene oor telkens dezelfde frequentie-inhoud had, en het geluid in het andere oor verschoven werd in frequentie over 1/6, 1/3 en 1 oct. Het resultaat was dat bij een toenemende frequentieverschuiving in één oor de JND stijgt (en de performantie dus daalt). Het was echter voor alle verschuivingen mogelijk nog ILDs waar te nemen. Voor een niet verschoven signaal was de JND in ILD ongeveer 2.5 dB. Bij verschuivingen van 1/6, 1/3 en 1 oct[13] nam deze toe met respectievelijk 0.5, 0.9 en 1.5 dB. De volledige studie wordt beschreven in hoofdstuk 3.

[13] Bij een verschuiving in frequentie van een octaaf (oct), verdubbelt de frequentie. Indien bijvoorbeeld 1000 Hz met 1 oct wordt verschoven, bekomen we 2000 Hz.

5 Perceptie van ILDs met bimodale stimulatie

We bepaalden de JND in ILD bij 10 gebruikers van een bimodaal systeem. In een eerste set van metingen gebruikten we signalen die een gelijkaardige toonhoogte teweeg brachten in beide oren, dit om de place mismatch te minimaliseren. In een tweede set was een aanzienlijke place mismatch aanwezig. De gemiddelde JND in ILD voor de gematchte signalen was 1.7 dB. Bij normaalhorenden worden in gelijkaardige tests JNDs in ILD van ongeveer 1 dB opgemeten, dus dit is een verrassend goed resultaat. Aangezien ILDs in realistische signalen kunnen oplopen tot 20 dB, volstaat de gevoeligheid van bimodale luisteraars om ILDs te gebruiken voor lokalisatie.

Uit de JND in ILD experimenten berekenden we ook luidheidsaangroeifuncties. Dit zijn functies die weergeven hoeveel luider het geluid aan het ene oor wordt voor een bepaalde verhoging van luidheid aan het andere oor. We vonden dat voor alle proefpersonen de luidheidsaangroeifuncties lineair waren op een decibel versus microampère schaal, maar dat de helling ervan afhankelijk was van het dynamisch bereik van beide oren. Bijgevolg is het mogelijk om een CI-spraakprocessor en een HA zo in te stellen dat ILDs niet verstoord worden door een verschil in luidheidsaangroei.

De volledige studie wordt beschreven in hoofdstuk 4.

6 Versterking van ILDs in de lage frequenties

Hoewel we aantoonden dat ondanks place mismatch ILDs nog waargenomen kunnen worden (zie sectie 4) en dat bimodale luisteraars gevoelig zijn voor ILDs (zie sectie 5), is er nog steeds een probleem. ILDs zijn namelijk fysisch enkel aanwezig in de hoge frequenties, terwijl het restgehoor van de meeste bimodale luisteraars beperkt is tot de lage frequenties.

Een ander probleem voor lokalisatie is dat het met klinische bimodale systemen niet mogelijk is om ITD-cues waar te nemen (zie sectie 2.4). Daarom gingen we na wat het effect is op het lokalisatievermogen als er geen ITDs aanwezig zijn. De conclusie is dat onder bepaalde voorwaarden het lokalisatievermogen even goed kan zijn met enkel ILD-informatie als met enkel ITD-informatie (zie hoofdstuk 5).

Om gebruikers van een bimodaal systeem toegang te geven tot ILD informatie in realistische signalen, ontwikkelden we een algoritme dat ILDs bepaalt aan de hand van de microfoonsignalen van beide oren en die dan introduceert in de lage frequenties van het akoestisch signaal (zie hoofd-

stuk 5). Via simulaties met normaalhorenden vonden we dat na versterking van ILDs de score op een lokalisatietest gemiddeld met 14° RMS fout[14] verbeterde.

7 Perceptie van ITDs met bimodale stimulatie

In een laatste studie gingen we de gevoeligheid na van bimodale luisteraars voor ITDs. Waar we bij ILDs gevoeligheid konden verwachten, lag dit bij ITDs anders omdat het neurale mechanisme voor detectie van ITDs gebruik maakt van de correlatie tussen beide oren, die bij bimodale stimulatie in het algemeen waarschijnlijk een stuk lager is dan bij akoestische of bilaterale CI-stimulatie.

We testten 8 gebruikers van een bimodaal systeem en 4 van hen waren gevoelig voor ITD met gemiddelde JNDs in ITD rond de $100 - 200\,\mu s$. Dit is vergelijkbaar met de JND in ITD bij gebruikers van bilaterale CIs, maar een stuk slechter dan bij normaalhorenden, waarbij JNDs tot $10\,\mu s$ opgemeten worden. ITDs in realistische signalen variëren van $0\,\mu s$ voor een signaal recht voor tot ongeveer $700\,\mu s$ voor een signaal volledig aan de linker- of rechterkant van het hoofd. Daarom volstaat de gevoeligheid van deze 4 proefpersonen om bruikbaar te zijn voor lokalisatie.

Er was een relatie tussen het al dan niet kunnen waarnemen van ITDs en het restgehoor. De 4 proefpersonen met ITD gevoeligheid hadden een gemiddelde gehoordrempel op 1000 en 2000 Hz van minder dan 100 dB SPL, terwijl die bij de andere 4 proefpersonen hoger was dan 100 dB SPL.

We vonden ook dat voor simultane bilaterale stimulatie van de gehoorzenuw het elektrische signaal gemiddeld vertraagd moet worden met 1.5 ms, ter compensatie van de tijd die de geluidsgolf nodig heeft om zich van het buitenoor tot in de cochlea te verplaatsen.

De volledige studie wordt beschreven in hoofdstuk 6.

8 Besluit

Gebruikers van een bimodaal systeem zijn gevoelig voor zowel ILD- als ITD-cues. Hun gevoeligheid is goed genoeg om realistische ILD- en ITD-cues te gebruiken voor lokalisatie. Door toepassing van een algoritme

[14] De RMS fout is een maat voor de fout die een proefpersoon maakt in een lokalisatietest. Hij wordt berekend aan de hand van het verschil tussen de werkelijke lokatie van een geluidsbron en de door de proefpersoon waargenomen lokatie. De RMS fout wordt hier uitgedrukt in graden en wordt gedefinieerd in sectie 1.6.2

dat ILDs versterkt, verbeterde het lokalisatievermogen in simulaties met normaalhorenden.

De signaalverwerking en instelling van huidige CI-spraakprocessoren en HAs moeten aangepast worden, zodat (1) de binaurale luidheidsaangroei lineair is, (2) beide apparaten gesynchroniseerd zijn, waarbij het elektrisch signaal met 1.5 ms extra vertraagd wordt en (3) de place mismatch minimaal is.

Chapter 1

Introduction

1.1 Motivation

Due to the success of cochlear implants (CIs), an increasing number of patients with residual hearing is implanted. When electric stimulation is combined with acoustic stimulation, the setup is called bimodal stimulation or electric acoustic stimulation (EAS). There can be residual hearing in either the implanted or non-implanted ear. If the acoustic stimulation is in the non-implanted ear (contralateral), we are dealing with a bilateral system. As localization of sound sources in normal hearing (NH) subjects is based on binaural cues (interaural time differences (ITDs) and interaural level differences (ILDs)), we can expect localization performance for users of a bilateral bimodal system to be better than for users of only a single CI. Indeed, it has been shown that localization performance improves in many subjects when fitting a contralateral hearing aid. However, performance is still poor compared to NH listeners (Ching et al., 2007).

While there are different technical reasons for this deficiency, even if they were resolved, it was not known before whether bimodal listeners are sensitive to the basic binaural cues. Therefore, in this thesis, the sensitivity of users of bilateral bimodal systems to binaural localization cues is assessed and a signal processing algorithm is proposed to improve their localization performance.

In this introduction, we will give a broad overview of hearing loss (section 1.2), hearing aids (section 1.3), cochlear implants (section 1.4), bimodal stimulation (section 1.5) and localization in NH and CI users (section 1.6), followed by an overview of the entire thesis (section 1.7). More specific introductory information is given at the beginning of each chapter.

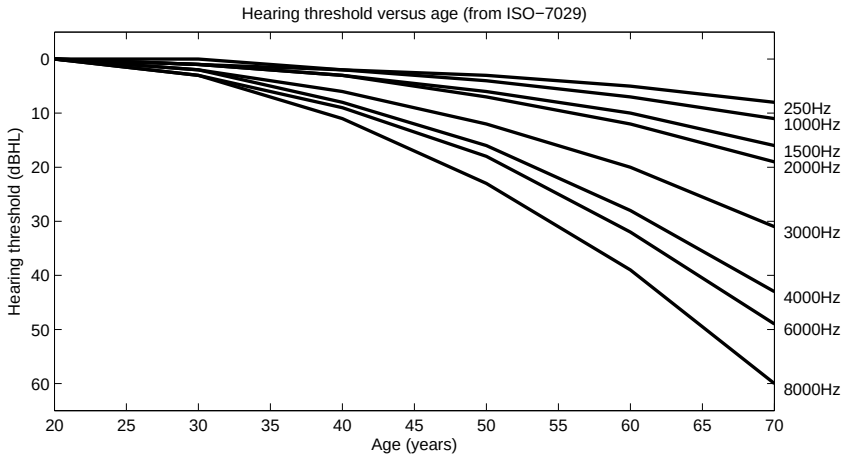


Figure 1.1: Median hearing thresholds versus age for otologically normal male subjects (from ISO-7029)

1.2 Hearing loss

Hearing loss is a common handicap. According to 2005 estimates by the World Health Organization (WHO), 278 million people worldwide have moderate to profound hearing loss in both ears. One in thousand newborns is affected by a severe hearing loss, either congenital or acquired. Moreover, the prevalence of hearing loss increases monotonically with age as hearing is irreversibly affected by noise induced trauma and age-related hair-cell degeneration. As a result, about half of the people aged 65 or older suffer from a mild to severe hearing loss.

In the ISO-7029 standard, quantiles of hearing thresholds for the otological normal population are given. The median values for male subjects are shown in figure 1.1. This indicates that a large part of the population will during their lifetime be confronted with high frequency hearing loss.

There are many causes of deafness. It can be inherited: if one or both parents or a relative are born deaf, there is a higher risk that a child will be born deaf. Hearing impairment may also be caused before or during birth for several reasons, including premature birth, infections during pregnancy[1] and the use of ototoxic drugs (Bauman, 2004). Later in life it

[1]Position Statements from the Joint Committee on Infant Hearing
`http://www.jcih.org/posstatemts.htm`

can be caused by infectious diseases such as meningitis, measles, mumps and chronic ear infections. Prolonged exposure to excessive noise, including working with noisy machinery, exposure to loud music (e.g., from an MP3 player (Fligor and Cox, 2004; LePage and Murray, 1998; Williams, 2005)) or other loud noises, can damage the inner ear and ultimately cause deafness.

While hearing impairment is easily concealed, it can be a severe handicap. Severely hearing impaired people often find themselves excluded from the NH society, because of their problems communicating with other people. This lack of human interaction can lead to many other problems, such as depression. Moreover, not only communication is problematic, also in other situations, such as traffic, hearing impairment can lead to accidents.

There are different degrees of hearing loss. Based on the pure tone average (PTA)[2] the following categories are distinguished: mild (PTA of $25 - 40\,\text{dB HL}$), moderate ($40 - 55\,\text{dB HL}$), moderate-severe ($55 - 70\,\text{dB HL}$), severe ($70 - 90\,\text{dB HL}$) and profound ($> 90\,\text{dB HL}$). An ear with profound hearing loss is also referred to as deaf. In this thesis we will mainly deal with subjects with severe to profound hearing loss.

In the current state of medical science, deafness cannot be cured. There do, however, exist assistive devices to facilitate communication and other aspects of hearing. The two main categories of assistive devices are hearing aids (see section 1.3) and cochlear implants (see section 1.4).

1.3 Hearing aids

Hearing aids exist in different shapes. Some types are placed behind the ear (BTE), some in the ear canal (ITC) and some even completely in the ear canal (CIC) (Dillon, 2001, Ch. 10). A hearing aid (HA) in its simplest form is an amplifier. However, modern HAs contain at least automatic gain control (AGC) and in most cases many other types of signal processing.

One major problem hearing impaired people have to cope with is reduced audibility. Soft sounds are not perceived anymore. This can be solved in a HA by amplifying all incoming sound. This, however, introduces another problem, namely that loud sounds can become too loud. In sensorineural hearing impairment, the threshold of loudness discomfort does not shift with the hearing threshold, such that the dynamic range is

[2] The pure tone average (PTA) is the average pure tone threshold at 500, 1000 and 2000 Hz

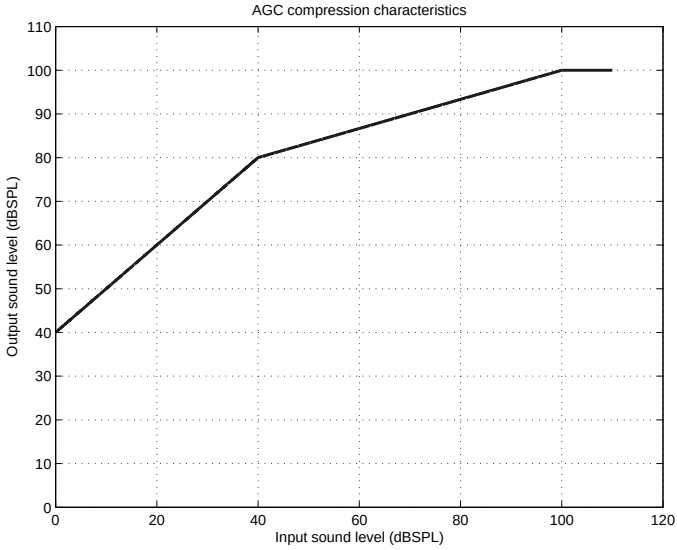


Figure 1.2: I/O characteristics of an example AGC. In this figure a wide dynamic range compression characteristic is shown. The first part is linear, followed by a non-linear part from input level 40 dB SPL with compression ratio 3, and output clipping at levels higher than 100 dB SPL.

reduced (Dillon, 2001, Ch. 6). In case of severe to profound hearing loss, the dynamic range can be reduced to as little as 10 dB.

To circumvent this problem, hearing aids reduce the dynamic range of the signal by the use of AGC. This is a signal processing block that monitors the *average* sound level and adapts the gain of the amplifier accordingly. It will use a larger gain for soft sounds than for loud sounds. An example HA compression characteristic is shown in figure 1.2. To limit signal distortion, AGC does not operate instantaneously, but uses time constants. The average level of the sound is considered and the gain is adapted after a certain interval of time. The attack time is the time that it takes to reduce the gain for loud input sounds and the release time is the time that is takes to increase the gain for soft input sounds to within 2 dB of the stationary level (IEC 60118-2). Typical attack times are in the order of 5 ms and release times are often longer than 20 ms (Dillon, 2001, Ch. 6).

As HAs rely on the function of the external, middle and inner ear, they are not useful for deaf subjects.

1.4 Cochlear implants

A cochlear implant (CI) is a device that bypasses a nonfunctional inner ear and stimulates the auditory nerve with patterns of electric current such that speech and other sounds can be perceived by profoundly deaf people. It is estimated that worldwide there are more than 152000 cochlear implant users. According to data collected by EuroCIU[3], in 2007 the total number of CI users in Europe was more than 47000, of which approximately 1000 are bilaterally implanted users and more than 22000 are children. In Belgium alone there were 1620 CI users. There are three main manufacturers active in the CI market: Cochlear, Advanced Bionics and Med-El. With its 65-70% market share, Cochlear is the largest player worldwide.

For the detailed operation of CIs, we refer to the literature (Clark, 2003; Zeng et al., 2004). In the next paragraphs we will give a short overview of the main components and discuss some design issues that are relevant for the remainder of this thesis.

Current CIs consist of two parts: an internal part that is implanted during a surgical procedure and an external part that is placed behind the ear and looks like a (rather big) hearing aid. Both parts communicate via a wireless link. An overview is shown in figure 1.3.

The main components of the internal part are a receiving coil, a decoder and stimulator and an electrode array. The receiving coil is placed on the skull, underneath the skin and connected to the decoder and stimulator. The stimulator is connected to an electrode array, inserted into the scala tympani of the cochlea from the base towards the apex. Current electrode arrays of Cochlear, Advanced Bionics and Med-El consist of respectively 12, 16 and 22 intracochlear electrodes. They are spaced equidistantly and a full electrode array insertion requires an insertion depth of 25 to 30 mm. Next to the intracochlear electrodes there are return electrodes which are implanted outside of the cochlea. A drawing of an electrode array implanted in a cochlea is shown in figure 1.4. The power necessary for electric stimulation is provided by the external part via the wireless link.

[3] EuroCIU, the European association of CI users, annually collects demographic data on the number of CI users. More information can be found on their website at `http://www.eurociu.org`.

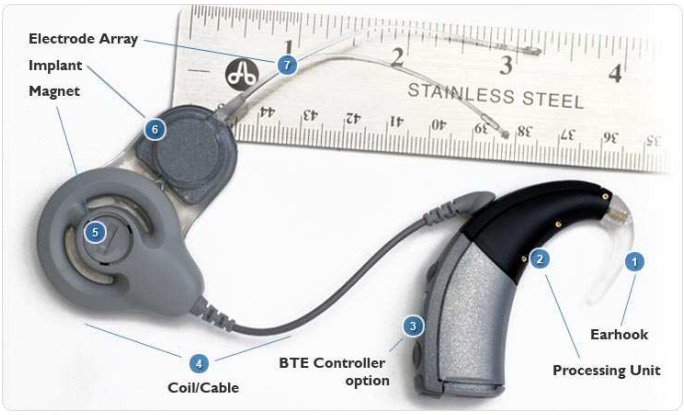


Figure 1.3: General overview of the internal and external parts of a cochlear implant system (reprinted with permission from Cochlear)

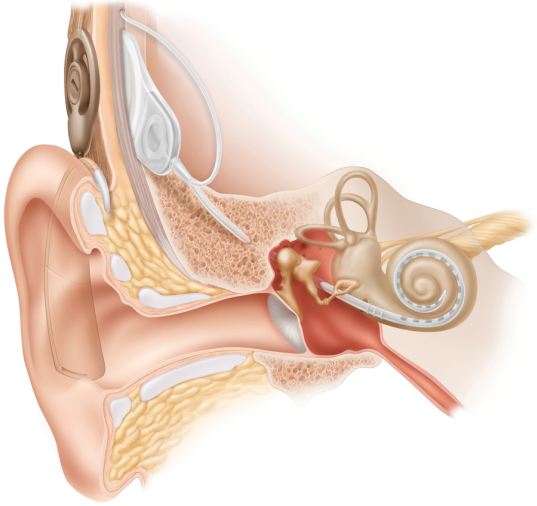


Figure 1.4: Drawing of an electrode array implanted in a cochlea (reprinted with permission from Cochlear)

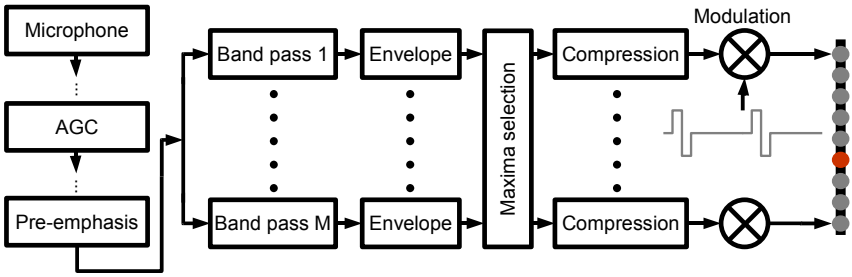


Figure 1.5: Block diagram of the signal processing in the ACE strategy

1.4.1 The speech processor

The external part, the so-called speech processor, converts the sound signal picked up by the microphone(s) to electric stimuli to be delivered on the implanted electrodes. There are many different speech processing strategies and most of them are aimed at speech understanding in quiet or in noise. In what follows, we will describe the N-of-M strategy, which is currently used by most CI users. We will focus on the ACE strategy, which is the implementation of the N-of-M strategy as used in the Cochlear Nucleus CIs.

A block diagram of the ACE strategy is shown in figure 1.5. After pre-emphasis, AGC and eventual other front-end processing, the signal from the microphone is sent through a filter bank. In each channel envelope detection is performed between the threshold (T) and most comfortable (C) stimulation levels. Then the N largest outputs are selected from the M channels of the filter bank, and logarithmic compression is performed according to the subject's personal settings. Stimulation patterns are generated by modulating electric current pulse trains with the resulting signals. These patterns are encoded and sent to the electrodes of the internal part via the wireless link. The number of signal processing channels is usually equal to the number of electrodes and there is a fixed mapping between channels and electrodes.

In the Nucleus system, amounts of current are commonly expressed in Current Units (CU)[4]. CU can be converted to μA by

$$I = 10e^{\frac{CU*\log(175)}{255}} \tag{1.1}$$

with I the electric current in μA and CU the number of current units.

[4]Current units are also called Clinical Units

In the entire signal processing of the speech processor, there are many parameters that have to be set individually per patient. The process of setting these parameters is called "fitting" or "mapping". During fitting, a MAP (set of individual parameter settings) is created by the audiologist. Most speech processors can hold more than one MAP and the subject can choose between MAPs using a switch on the speech processor. The main parameters in a MAP are

electrodes Each electrode can be disabled in case of a malfunction or unwanted side-effects such as stimulation of the facial nerve or according to specific signal processing schemes.

mode of stimulation (electrode configuration) the reference electrode to be used, can either be one or both extracochlear electrodes (so-called monopolar mode) or one or more intracochlear electrodes (so-called bipolar mode).

T-levels The threshold level is the smallest current that elicits an auditory sensation for each electrode.

C-levels The comfort level is the largest current that results in a comfortably loud auditory sensation for each electrode.

Q-factor The shape of the I/O function of the logarithmic map is influenced by the Q-factor. An example of compression characteristics for different Q factors is shown in figure 1.6.

pulse rate Most speech processors use a fixed pulse rate per electrode.

front-end processing Most front-end processing such as noise reduction and automatic gain control can be enabled or disabled and parameters can be set.

1.4.2 Design issues

The design of a speech processing strategy is a complex process that requires many iterations between algorithm development, psychophysical tests and take-home experiments. In the following sections we will highlight a few aspects of the speech processing strategy that are of particular importance for the remainder of this thesis.

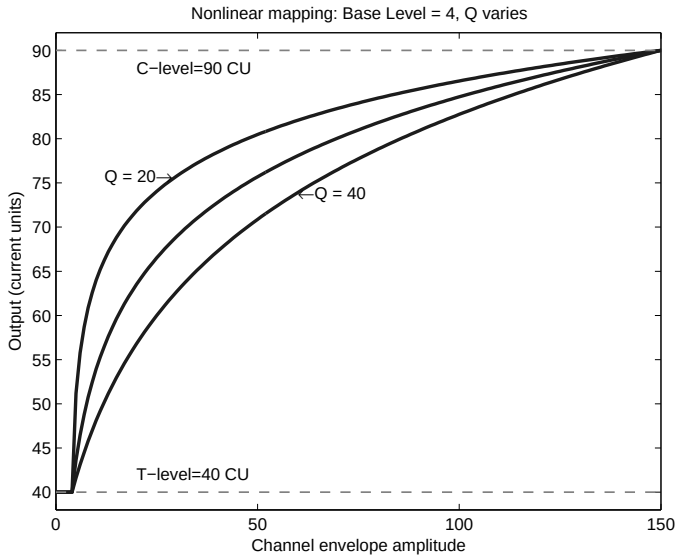


Figure 1.6: CI compression characteristics for different values of Q

The filter bank

Many types of filter banks have been suggested for use with speech processors, each with specific advantages and disadvantages. An important property of the filter bank is that it indirectly associates acoustic frequency ranges with electrode locations in the cochlea. The cochlea has a tonotopic organization, i.e., specific locations in the cochlea are stimulated by specific frequency components of the acoustic signal (Greenwood, 1990). Therefore, specific electrodes also correspond with specific frequency ranges. If the cutoff frequencies of the filter bank do not correspond to the frequency ranges normally associated with the places in the cochlea that are stimulated by the electrodes, the recipient needs a longer period of adaptation before speech perception reaches maximum performance (Fu et al., 2002; Rosen et al., 1999).

Loudness perception

Loudness perception using a CI is governed by perceptual factors and by 3 main technical factors: the AGC, logarithmic mapping and conversion of current units (see below) to μA. The AGC operates in the same way as

in a HA (see section 1.3) and is more or less linear on a short time scale. In contrast, the logarithmic mapping is instantaneous (and therefore obviously non-linear) and its I/O function depends on the T-levels, C-levels and Q-factor set during fitting (see figure 1.6). The results after logarithmic mapping are values in so-called "current units". These units are converted to μA in a non-linear way by the (implanted) stimulator. The combination of logarithmic mapping and conversion from current units is supposed to model the compression in the NH cochlea. Its parameters are however determined by optimization of speech perception performance (Fu and Shannon, 1998) and not by optimization of linearity of loudness perception or correspondence with NH. Therefore, the entire loudness processing chain will in many cases be non-linear and different from NH.

Bilateral implantation

Because of the high cost of cochlear implantation, most recipients are implanted unilaterally. While adding a second CI does not provide the same amount of benefit as adding the first one, there are important advantages associated with bilateral implantation, such as improved localization performance and improved speech perception in noise (Ching et al., 2007).

1.4.3 Subject performance with CIs

The design of current CIs is a crude approximation of the normal peripheral auditory processing and there is a large variation in performance over different patients. Nevertheless, in the average recipient, CIs function surprisingly well.

One of the main deficiencies of a CI is the severely reduced spectral resolution, compared to NH. Using noise band vocoder simulations, Dorman et al. (1997) and Shannon et al. (1995) assessed speech perception performance in quiet in NH subjects as a function of the number of channels. Performance increased when increasing the number of channels from one up to about six and remained stable thereafter. Similar results were obtained for CI users (Zeng et al., 2004, chap. 8). Therefore six channels may be considered sufficient for speech perception in quiet.

In noise the situation is different. Friesen et al. (2001) showed that for speech perception in noise, performance improved for NH subjects when increasing the number of channels up to 20. In CI subjects, however, performance was asymptotic at 7 to 10 channels. This suggests that the CI subjects could not make use of more than 7 to 10 discrete channels.

Van Wieringen and Wouters (2008) presented the speech perception results in quiet and in noise for 16 CI users using their clinical processors. They showed that speech perception performance for the LIST sentences in quiet was for most subjects nearly 100% correct. However, speech perception performance in noise was more variable between subjects and much worse than in NH subjects. The best CI subjects acquired a speech reception threshold (SRT) between 0 and 5 dB for the LIST sentences, while for NH subjects the average SRT was −7.8 dB, which is a huge difference considering performance in daily life.

1.5 Bimodal stimulation

The conventional CI candidate is profoundly deaf. However, due to the success of CIs, many CI users perform better than some severely hearing impaired HA users. Therefore, more and more subjects with residual hearing are being implanted. This gives rise to a new population of CI users who have residual hearing in either the implanted (ipsilateral) or the non-implanted ear (contralateral). In most cases residual hearing is only present at low frequencies (up to 1000 Hz).

Surgical techniques for preservation of residual hearing in the implanted ear (Gstoettner et al., 2004, 2006; Kiefer et al., 2005) and a special short electrode array have been developed that interfere less with low frequency residual hearing (Gantz and Turner, 2004; Turner et al., 2004). In this thesis, we will, however, focus on acoustic stimulation in the non-implanted (contralateral) ear. Whenever the term "bimodal stimulation" is used, *bilateral* bimodal stimulation is meant.

The addition of acoustic stimulation via a HA to a CI has been shown to slightly improve speech recognition performance in quiet and greatly improve speech recognition performance in noise with a competing talker (Ching et al., 2007; Dorman et al., 2007a; Kong and Carlyon, 2007; Kong et al., 2005; Turner et al., 2004; Tyler et al., 2002).

While it is evident that the combination of a CI and a contralateral HA offers many advantages, there are several technical problems. A clinical bimodal hearing system currently consists of a CI and a HA which are designed separately and are in many cases also fitted separately. This leads to discrepancies between the ears. In the following paragraphs we will describe four different problems: binaural loudness growth, synchronization, place mismatch and sound quality.

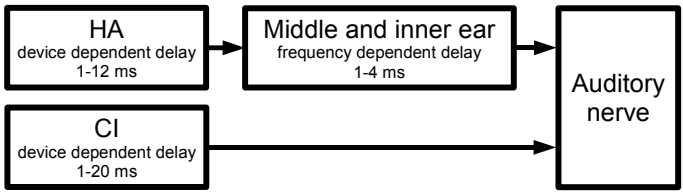


Figure 1.7: Illustration of synchronization problems in bimodal systems

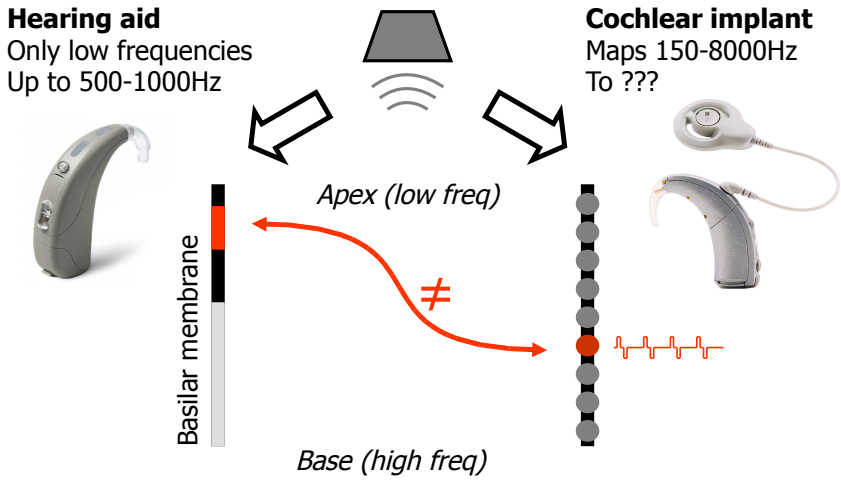


Figure 1.8: Illustration of place mismatch between electric and acoustic stimulation

1.5.1 Problems with current clinical bimodal systems

Binaural loudness growth

A first problem with current bimodal systems is non-linear binaural loudness growth. The CI and HA both contain AGCs that have different parameters and operate independently of each other, leading to uncontrolled binaural loudness balance. Moreover, the CI contains compression (see section 1.4.2), which is not necessarily the same as the compression that occurs in the other severely impaired ear.

Synchronization

A second problem is the synchronization of the CI and HA (see figure 1.7). In the electric path, there is a device dependent but fixed processing delay of $10 - 20$ ms. In the acoustic path there is, on the one hand, a device dependent (and sometimes even frequency dependent) processing delay of the HA and on the other hand a frequency dependent delay of the sound wave traveling through the middle and inner ear. In most cases the total delay of the acoustic and electric path will not be the same, leading to problems perceiving binaural cues.

Place mismatch

A third problem is the place mismatch between electric and acoustic stimulation (see figure 1.8). The acoustic signal is – according to its frequency content – presented at a certain place in the cochlea[5]. The electric signal is sent to a certain electrode, whose place in the cochlea will in many cases not correspond to the place of stimulation in the other ear. This is due to, on the one hand, the use of a filter bank in the CI speech processor which is not customized per patient (see section 1.4.2) and, on the other hand, the limited amount of residual hearing. Higher frequencies that are stimulated by the CI cannot be perceived with the residual hearing. In section 1.5.2 some methods for matching the place of excitation are reviewed.

Sound quality

Finally, a fourth problem is differences in sound quality between the ears. While this is a very subjective issue, it is clear that electric stimulation yields a percept that is in many cases very different from that for acoustic stimulation (McDermott and Sucher, 2006). While subjects adapt to these differences, they are uncomfortable for the subject, may be detrimental for integration of sound between ears and an indication that the CI signal processing should be changed such that the electric signal is perceptually more similar to the acoustic signal.

[5] In NH subjects, there is a fixed correspondence between acoustic frequencies and places in the cochlea. However, certain hearing impairments can lead to a shift in the frequency-place mapping (Moore, 1995).

1.5.2 Matching the place of excitation

Different approaches have been suggested for matching the place of excitation between an acoustically and electrically stimulated cochlea. The most straightforward approach is pitch matching (Baumann and Nobbe, 2006; Blamey et al., 1996; Boex et al., 2006; Dorman et al., 1994, 2007b). Both the electric stimulation rate and place affect the perceived (matched) pitch (Blamey et al., 1996). However, as with high pulse rates the temporal pitch percept saturates (Shannon, 1983; Zeng et al., 2004) and the pitch only varies with electrode location, it is hypothesized that stimuli that elicit a similar pitch percept stimulate the same place in the cochleas.

Boex et al. (2006) measured the acoustic pitch corresponding to the place pitch elicited by stimulation of certain electrodes of the cochlear implant in 6 users of bimodal systems. For the most apical electrode of each subject, they found pitches of 460, 100, 290, 260, 570 and 300 Hz. These pitches are lower than would be expected based on Greenwood's function (Greenwood, 1990).

Dorman et al. (2007b) compared computerized tomography (CT) scans and pitch matching data from a single subject with a Med-El Combi 40+ CI. They found that for insertion angles greater than 450 degrees or greater than approximately 20 mm insertion depth, pitch did not decrease beyond approximately 420 Hz. From 20 to 15 mm insertion depth pitch estimates were about one-half octave lower than predicted by the Greenwood function. From 13 to 3 mm insertion depth the pitch estimates were approximately one octave lower than predicted by the Greenwood function. The pitch matches for electrodes 1-11 were respectively 441, 404, 397, 495, 666, 927, 1065, 1230, 1550, 2584, and 3449 Hz.

The problem with the pitch matching method is that pitch perception using the CI might change over time for a certain period after first switching on the CI (Reiss et al., 2007). The result is that a stimulus that is at one time perceptually the same in both ears, may not be the same any more at another time (e.g., a few months later). Therefore, as electrode locations in the cochlea are fixed, the location in the electrically stimulated cochlea may not correspond with the pitch matched location in the acoustically stimulated cochlea.

Another method for matching the place of excitation is contralateral masking (James et al., 2001). For a fixed location in one cochlea, the amount of masking by a contralateral stimulus at several locations in the other cochlea is determined. It is assumed that the stimulus with the greatest masking power is tonotopically most similar. However, this procedure is very time consuming and does not yield a very precise result.

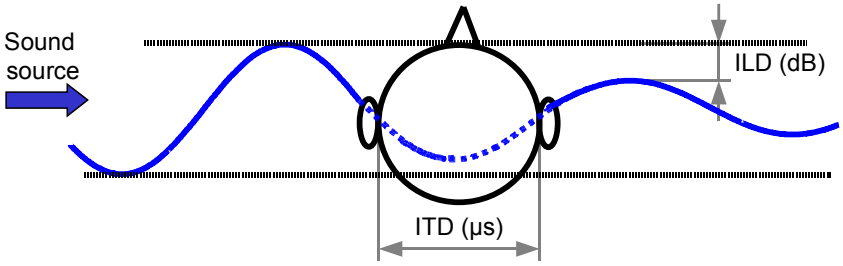


Figure 1.9: Illustration of ILD and ITD for a sound incident from the left side of the head (-90°)

In chapter 6 we suggest a novel method for matching the place of excitation using sensitivity to ITDs.

1.6 Localization

Humans can localize sound sources in the left-right direction, but also in the front-back and above-below direction. While monaural spectral cues are used for localization in the front-back and above-below direction, binaural cues are used for localization in the left-right direction (the frontal horizontal plane). We will focus on binaural sound localization in the horizontal plane. This is the plane parallel with the ground.

Part of the sound source localization process in NH persons was already understood more than 120 years ago. As part of his famous duplex theory, Strutt (1877) observed that if a sound source is to the left of a listener's forward direction, an acoustic shadow will be cast by the head over the right ear, causing the signal at the right ear to be lower in level than the one at the left ear (see figure 1.9). The resulting interaural level difference (ILD) can be used to localize the sound. Similarly, due to the limited speed of sound, the waveform will arrive earlier at the left ear than at the right ear (see figure 1.9). The resulting interaural time difference (ITD) can be used to localize the sound. ILDs and ITDs are still considered the basic cues for localization of sound sources in the horizontal plane. Reviews of sound localization can be found in Akeroyd (2006); Blauert (1997); Hartmann (1999); Moore (1995).

In the next sections, we will first review some methods to assess localization performance and then focus on the basic cues and how well they are perceived by either NH subjects, users of bilateral CIs or users of a

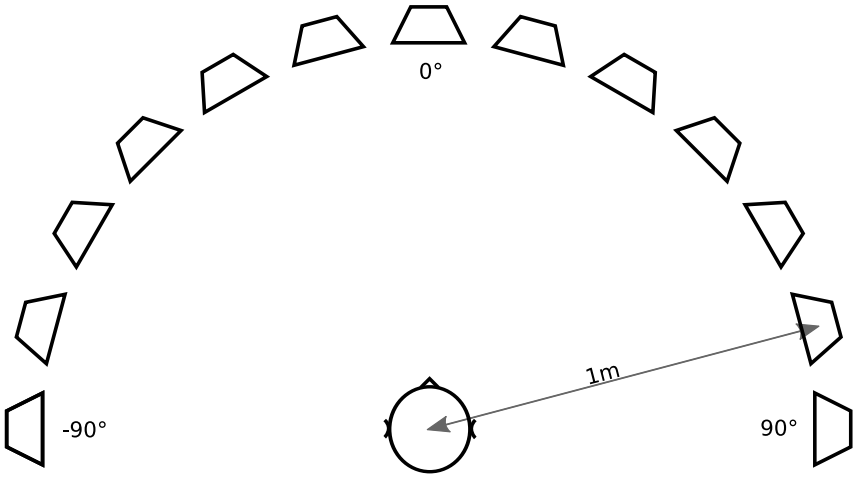


Figure 1.10: Schematic of the localization test setup at ExpORL

bilateral bimodal system.

1.6.1 Measuring localization performance

There are many ways to investigate sound source localization. Questionnaires can be used to assess localization in daily life, subjects can be tested in a laboratory setup with loudspeakers or signals can be presented via headphones to assess sensitivity to the basic cues separately.

Measuring the localization error

A straightforward method to measure localization performance is to measure the difference between real and perceived stimulus location. In this method, the subject is typically seated in the middle of an array of loudspeakers, a sound is played from one of the loudspeakers and the subject is asked to indicate where the sound came from. In our lab we use an array of 13 loudspeakers, spaced by 15° at a distance of approximately 1 m from the listener (see figure 1.10). In such a setup typically a localization error measure such as the root mean square (RMS) error or absolute error between source location and subject response in degrees is calculated. We define the direction in front of the subject as 0°, the right hand side as 90° and the left hand side as -90°. The location exactly at the back of the subject is at 180°.

Many different error measures are used in localization experiments. In this thesis, localization error will mostly be reported as the RMS localization error

$$E_{RMS} = \sqrt{\sum_{i=1}^{N} \frac{(S_i - R_i)^2}{N}} \tag{1.2}$$

with S_i the location of the i-th stimulus (in degrees), R_i the location of the i-th response and N the number of presentations.

Measuring the just noticeable difference

Other methods measure the resolution of different aspects of the localization system. The resolution is commonly expressed as a just noticeable difference (JND): the smallest difference of a certain parameter that can be discriminated. Smaller values of the JND indicate better performance. The JND can be determined in a discrimination task or in a pure lateralization task. JNDs can for example be determined in angle, ILD or ITD (cf following sections). In a discrimination task, in each trial two stimuli have to be discriminated. For example, in each trial two stimuli can be presented from a different angle and the subject has to respond whether the second one was on the left or right hand side of the first one. In a pure lateralization task, only one stimulus is presented per trial and the subject has to indicate whether it was on the left or right hand side of the middle.

Sensitivity to binaural cues is in most cases highest around the value corresponding to the location right-in-front (0°). Therefore, in most discrimination tasks the right-in-front value is used as the reference signal (e.g., ILD=0 dB or ITD=0 µs).

The function relating the value of a variable (e.g., ILD in dB or ITD in µs) to performance (e.g., in % correct) is called a performance-intensity function or a psychometric function. JNDs can be measured using adaptive or constant stimuli procedures, but either case comes down to estimating the value of the variable at a certain performance level. In a constant stimuli procedure the psychometric function is determined and the JND can be defined as the value of the variable at the point halfway between the chance level and theoretical best performance (in a two alternative forced choice task, this would correspond to the 75% correct point[6]). An

[6]The 75% correct point, is the value of the variable for which the subjects answer correctly in 75% of the cases. If, for example, the minimum audible angle (MAA) is

example psychometric function for the determination of the JND in ITD is shown in figure 6.4 on p143.

When measuring sensitivities, it is important to take care that only those cues are measured that are under investigation. If a non-desired cue cannot be eliminated, it is typically roved[7] such that, on average, the results are not influenced. In a discrimination task, one has to take care that the subject does not use information from a previous trial to respond (Hartmann and Rakerd, 1989).

Measuring the JND in angle or minimum audible angle

The resolution measure that is closest to "real" localization, is the JND in angle or the minimum audible angle (MAA). This is the smallest angle that can be discriminated, or the JND in angle. In a discrimination task it can be measured by playing a sound from one of two different loudspeakers right in front of the subject and asking which one was playing. This can be done for different angles between the two speakers, such that a psychometric function can be determined from which 75% correct point is derived. Note that the MAA is a *relative* localization measure, while the localization error (see section 1.6.2) is an *absolute* localization measure.

Measuring the just noticeable difference in ILD and ITD

To assess sensitivity to individual localization cues, it is necessary to manipulate them individually. This is not always possible in a free field setup. Therefore, headphones or other means of direct stimulation such as insert phones or computer interfaces to CIs are frequently used in this kind of experiment. When artificial signals under headphones are used, the process of perceiving them at either side of the head is called lateralization rather than localization (Plenge, 1974). The JND in ILD or ITD can be measured as described above in section "The just noticeable difference".

In the next sections, results of measurements of JNDs to the different cues will be given for NH subjects, users of bilateral CIs and users of a bilateral bimodal system.

---

measured, the variable is the angle between the speakers and the 75% correct point will be expressed in degrees.

[7] Roving a cue involves setting it to a different value in every trial. For example, loudness roving is used to eliminate undesired loudness cues.

1.6.2 Localization error

In the localization test setup at ExpORL, if a stimulus is presented three times from each of the 13 different loudspeakers in our setup, the chance level is 76.4° RMS error. For NH subjects the mean localization error ranges between 6.8° and 21.3°, depending on the stimulus (Van den Bogaert et al., 2006). More data for NH subjects with different stimuli are presented in chapter 5.

While slightly better than chance level, localization using a single CI is poor. Grantham et al. (2007b) report adjusted constant errors around chance level for three CI subjects and around 40° for three other CI subjects, whereas NH listeners obtain an average score of 5.6° in this setup. In three different studies (Ching et al., 2004; Dunn et al., 2005; Seeber et al., 2004), when fitting a contralateral HA, localization performance improved, but not for all subjects and only slightly. As different test setups, fitting procedures and localization error measures are used in these studies, the results are hard to compare. Dunn et al. (2005) reported that 2 of 12 users of a CI and HA could localize sounds. RMS errors ranged from 27° up to 48°. Seeber et al. (2004) tested 11 subjects and reported that 1 subject showed very good localization performance, 2 performed above chance level, 4 could only discriminate the left and right side and 4 showed no localization ability at all. The 4 subjects with the best residual hearing performed best on the localization tasks. Ching et al. (2004) tested 18 adults and reported that 12 showed benefit with the addition of a contralateral HA and 6 did not. A review is given by Ching et al. (2007).

Across studies on bilateral cochlear implantation, 89% of adults show binaural advantages when using both devices (Ching et al., 2007). Again, differences between studies make it hard to compare the results.

1.6.3 Minimum audible angle

For NH subjects the MAA for sinusoidal stimuli is around 1° for sounds directly in front and increases up to 9° for sounds originating from the side of the head. It is lowest at low stimulus frequencies, and there is a region of inferior performance between 1500 and 1800 Hz. This is consistent with the duplex theory; Above 1500 Hz the ITD becomes ambiguous and up to 1800 Hz ILDs are small (Mills, 1958).

MAAs reported for adult bilateral CI users vary widely amongst studies, subjects and used devices. While for a few subjects reported MAAs are close to those found in NH, generally performance is much worse or even

unmeasurable (Grantham et al., 2007a; Nopp et al., 2004; Seeber et al., 2004; Senn et al., 2005; Van Hoesel et al., 2002; Verschuur et al., 2005).

There are to our knowledge no studies that measure the MAA in adult bilateral bimodal subjects. Litovsky et al. (2006) measured the MAA in eight bimodally stimulated children. Four of them obtained a clear bilateral benefit but absolute performance ranged from 11° to 72°.

1.6.4 ILD

The ILD is due to the head shadow effect, i.e., the effect of the head attenuating the sound arriving at one ear. Due to the acoustic properties of the head and pinnae, the ILD is strongly dependent on frequency. Also, the ILD increases with increasing frequency. This is caused by the fact that sound waves are diffracted if their wavelength is longer than the diameter of the head. For sound sources in the far field (i.e., further away than approximately 1 m), ILDs are considered useful for localization at frequencies higher than about 1500 Hz (Moore, 1995). In figure 5.6 on p120 ILDs are shown per frequency for different angles of incidence, as measured using a artificial head. It is clear that the magnitude of ILD cues increases with frequency and angle of incidence and that the distribution of ILDs over frequencies depends on the angle of incidence.

Mills (1960) presented 5 subjects with a reference with no ILD, followed by a stimulus with an ILD. The stimuli were pure tones. Using the method of constant stimuli, the JND in ILD was determined from half the interquartile separation of the psychometric curves for each subject. JNDs were around 1 dB for 1000 Hz, somewhat smaller for lower frequencies and around 0.5 dB for frequencies higher than 1000 Hz.

Yost and Dye (1988) measured JNDs in ILD for pure tones and different reference signals at 75 % correct, using a linear fit of the psychometric curve. For the reference at ILD = 0 dB JNDs were found of approximately 0.75, 0.85, 1.20, 0.70 and 0.73 dB for 200 Hz, 500 Hz, 1000 Hz, 2000 Hz and 5000 Hz. In the 2AFC procedure, subjects perceived one stimulus on the right side and one on the left side and had to respond which one was on the right.

While ILDs are small at low frequencies, the auditory system is nevertheless sensitive to ILDs over its entire frequency range. Low frequency ILD cues are especially used for localization of nearby sounds (Brungart, 1999; Brungart and Rabinowitz, 1999; Brungart et al., 1999). Summarizing, for NH subjects, the JND in ILD is relatively constant across a large range of frequencies and is in the range of $1 - 2$ dB (Feddersen et al., 1957; Mills, 1960; Yost and Dye, 1988). Performance is best if the reference is

around ILD=0 dB.

JNDs in ILD in bilateral CI users have been measured using the audio input of the Med-El implant. Senn et al. (2005) reported a JND of 1.2 dB difference in electric voltage at the audio input and Laback et al. (2004) reported JND values of 1.4 up to 5 dB . In the latter study, stimuli were chosen such that the pitch percept evoked by stimulation of the active electrodes at the two sides corresponded. Lawson et al. (1998) and van Hoesel and Tyler (2003) on the other hand used an experimental processor to directly stimulate the Nucleus implant, bypassing the clinical speech processor. Lawson et al. (1998) found JNDs of 1-4 current units, which equals $0.09 - 0.35$ dB change in electric current and van Hoesel and Tyler (2003) found JNDs of $< 0.17 - 0.68$ dB change in electric current.

JNDs in ILD for users of a bilateral bimodal system are assessed in chapter 4.

1.6.5 ITD

ITD cues are mainly functional at lower frequencies, because as soon as half of the wavelength of the sound equals the distance between the eardrums, the ITD becomes ambiguous. Moreover, phase locking of the auditory nerve fibers is only functional up to $4 - 5$ kHz (Moore, 2003). Therefore, for signals without envelope fluctuations, changes in ITD become undetectable above about 1500 Hz (Yost, 1974). ITDs range from 0 µs for sounds incident from 0° to around 700 µs for sounds incident from 90° (Kuhn, 1977). For low frequency stimuli, NH listeners are sensitive to ITDs with JNDs as low as 10 µs (Mills, 1958; Yost, 1974). Detection thresholds are again smallest when the reference is around 0°.

ITD cues are not only available in the fine structure of a signal, but also in the envelope of complex signals. This allows the binaural system to localize a high-frequency signal using ITD in the onset, offset or envelope (Bernstein and Trahiotis, 2002, 1985b; Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976). However, if both fine structure and envelope cues are available, the fine structure cues are dominant (Bernstein and Trahiotis, 1985a). JNDs in envelope ITD in modulated high frequency signals are in many studies found to be comparable to JNDs in ITD for low frequency stimuli (Bernstein and Trahiotis, 2002; Henning, 1974).

While there is quite some inter-subject variability when measuring sensitivity to ITDs in sinusoidal amplitude modulated signals, performance is seen to increase using so-called transposed signals. These are high frequency carriers modulated with a *half wave rectified* envelope. The half

wave rectification models the naturally occurring rectification in low frequency signals on the basilar membrane (Bernstein and Trahiotis, 2002, 2004, 2005, 2007). Bernstein and Trahiotis (2002) found that JNDs in ITD were smaller for transposed stimuli than for amplitude modulated stimuli and for low modulation frequencies ($< 128\,\mathrm{Hz}$) were even smaller than for their pure tone counterparts.

Recent studies have shown that users of bilateral CIs are sensitive to ITDs, although much less so than NH listeners. The best JNDs reported for pulse trains of about 100 pps are around $100 - 200\,\mu\mathrm{s}$ and for higher pulse rates JNDs are much higher or immeasurable (Laback et al., 2004; Lawson et al., 1998; Long et al., 2003; Majdak et al., 2006; Senn et al., 2005; van Hoesel, 2004, 2007; van Hoesel and Tyler, 2003).

The JND in ITD for users of a bilateral bimodal system is assessed in chapter 6.

1.6.6 Monaural and visual cues

In addition to the binaural ILD and ITD cues, there are also monaural cues that can be used for the localization of sound sources. The asymmetric shape of the pinnae introduces direction-dependent filtering of the incoming sound signal, which can be used to determine the source location if the spectrum of the sound source is known. If the level of a sound source is known, the head shadow effect can also be used monaurally.

Monaural cues are used for localization of sound sources from any direction, and are next to head movements the main cues available for perception of distance and elevation and for resolving front-back confusions. This is due to the fact that ILD and ITD cues are ambiguous within the so-called cone of confusion: for sound sources on the entire surface of a cone with the tip at the center of the head, overall ILD and ITD are the same.

In addition to monaural and binaural cues, visual cues are also taken into account when localizing sounds (Wallach, 1940). Therefore they should be considered when performing localization experiments (Lewald and Getzmann, 2006; Perrett and Noble, 1995). Examples of visual information influencing auditory perception are the McGurk effect (McGurk and MacDonald, 1976) and the ventriloquist effect (Bertelson, 1999, pp.347-362).

1.6.7 Head related transfer functions

The whole array of cues available for localization can be expressed using a so-called head related transfer function (HRTF). As they depend on the shape of the head and pinnae, HRTFs are strongly subject dependent and can be measured for a certain subject and a certain sound source location. Considering binaural HRTFs, the difference in phases indicates ITDs and the difference in amplitudes indicates ILDs.

The binaural system combines ILD and ITD to a certain extent of lateralization. This process is complex, depends on many factors and is not yet completely understood (Domnitz, 1973; Hafter and Carrier, 1972; Hafter et al., 1990; Palmer et al., 2007; Phillips and Hall, 2005; Phillips et al., 2006; Yost, 1981). As ILDs and ITDs both yield a percept of lateralization, they seem to be interchangeable, but that is not entirely the case (Hafter and Carrier, 1972).

1.6.8 Adaptation to changes in localization cues

The shape of the head and pinnae are strongly individual. As the brain relies on these shapes to localize sounds, they must be learned initially and must continuously be "calibrated" to cope with changes. The topic of learning sound localization cues is reviewed by Wright and Zhang (2006). Humans can adapt to changes in both ILD and ITD cues.

Hofman et al. (1998) artificially modified the shape of the pinnae using molds. They observed that immediately after the modification sound elevation localization was dramatically degraded. Later, performance steadily improved. Moreover, after the experiments, the subjects could localize accurately with both normal and modified pinnae cues. This indicates that the brain not only adapts to changes in HRTFs, but can also store several "templates".

Javer and Schwarz (1995) conducted a similar study for ITD cues. They required NH subjects to wear a HA and they inserted a fixed delay in one ear. The subjects gradually adapted significantly but not completely to the distortion over the course of several days. A few minutes after removal of the HA, localization was back to normal.

While different learning patterns occur, both ILD and ITD detection can be improved by training (Rowan and Lutman, 2007; Wright and Fitzgerald, 2001). While different conclusions are reached in different studies, overall there seems to be a pattern of greater modifiability of ILD than ITD processing.

1.7 Thesis outline

In this thesis, we assess sensitivity of users of bilateral bimodal systems to the basic localization cues (ILD and ITD) and suggest changes to the current CI and HA signal processing to improve localization performance. In the following paragraphs we give a chapter by chapter overview of the thesis.

To measure sensitivity to ILD and ITD, clinical HAs and CIs cannot be used because they influence these cues in an uncontrolled manner. Therefore an experimental platform is required that allows many psychophysical procedures and direct control of an acoustic transducer and a CI. We developed a generic platform for psychophysical experiments, called APEX 3. Its development and function are described in **chapter 2**. APEX 3 is used as an experimental platform in all subsequent chapters. The development of APEX 3 was published in Francart, van Wieringen, and Wouters (2008e). In the **appendix**, an APEX 3 module for automatic testing of speech perception is described, which was published in Francart et al. (2008c).

As explained in section 1.5 and figure 1.8, for users of a bilateral bimodal system, there is often a mismatch in place of excitation in the cochlea between the ears. In **chapter 3** it is assessed whether NH subjects can perceive ILDs when a frequency shift is introduced in one ear. For 4 different base frequencies the influence of a frequency shift of 1/6, 1/3 and 1 oct in one ear is assessed on the JND in ILD. The stimuli are uncorrelated 1/3 oct wide noise bands. The results presented in chapter 3 were published in Francart and Wouters (2007).

In **chapter 4** sensitivity to ILDs and loudness growth is measured for bilateral bimodal subjects. Two sets of experiments are done. In the first set the most apical electrode is used together with an acoustic signal that is matched in pitch. In the second set the most basal electrode is used to determine the effect of unmatched stimulation. Sensitivity to ILD was assessed by determining the JND in ILD in loudness balancing experiments. From these balancing experiments loudness growth functions between electric and acoustic stimulation were determined. The results presented in chapter 4 were published in Francart, Brokx, and Wouters (2008a).

ITD perception with clinical bimodal systems is not feasible in the short term. Therefore, in the first experiment of **chapter 5** we assessed whether it is possible to localize properly with only ILD cues by measuring localization performance of NH subjects under these circumstances. In chapter 4, it is shown that bimodal subjects are sensitive to ILD but they do not have

sufficient high frequency residual hearing to perceive real-world ILD cues. Therefore in the second experiment of chapter 5, the development and evaluation of an algorithm for automatic introduction of ILDs cues into the low frequencies are described. The results presented in chapter 5 are described in Francart, Van den Bogaert, Moonen, and Wouters (2008d).

Due to the place mismatch and synchronization issues described in section 1.5 and indicated by poor performance on localization tasks, users of clinical bimodal systems cannot perceive ITDs. Using our experimental setup, in **chapter 6** we assessed sensitivity to ITD of bimodal listeners. The results presented in chapter 6 were published in Francart, Brokx, and Wouters (2008b).

Finally, in **chapter 7** general conclusions are drawn and suggestions for further research are given.

Chapter 2

APEX 3: a multi-purpose test platform for auditory psychophysical experiments

To assess sensitivity to binaural cues with bilateral bimodal stimulation, a test platform is required with strict specifications concerning control over psychophysical procedures and stimuli presented to the subject. In this chapter both the hardware and software of the test platform are described.

Abstract

The hardware for bilateral bimodal stimulation consists of an experimental speech processor (L34) for electric stimulation and a multi channel sound card (RME Hammerfall) for acoustic stimulation. They are synchronized using a trigger signal. The software is a test platform for auditory behavioral experiments, which is called APEX 3. It provides a generic means of setting up experiments without any programming. The supported output devices include sound cards and cochlear implants from Cochlear Corporation and Advanced Bionics Corporation. Many psychophysical procedures are provided and there is an interface to add custom procedures. Plug-in interfaces are provided for data filters and external controllers. APEX 3 is supported under Linux and Windows.

In section 2.1, first the used hardware is described (section 2.1.1) and then the used software (section 2.1.2). The remainder of this chapter describes the APEX 3 software platform.

2.1 Introduction

A generic test platform was developed that allows many types of psychophysical procedures or speech perception experiments to be performed.

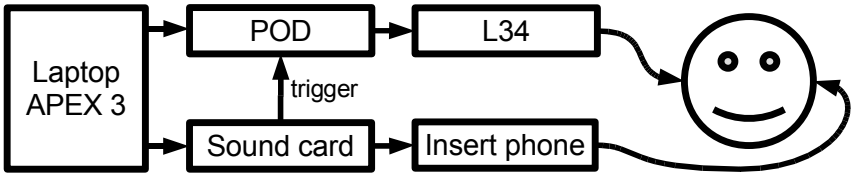


Figure 2.1: Experimental setup for synchronized electric acoustic stimulation

It can provide electrical stimulation via direct specification of pulse sequences and it can control the acoustic path via a sound card. Specific requirements for researching the perception of localization cues with bimodal stimulation are high acoustic output levels and control over the synchronization between electric and acoustic stimulation. An overview of the entire system is shown in figure 2.1.

2.1.1 The hardware platform

For acoustic stimulation we selected an RME Hammerfall DSP sound card connected to an insert phone of type Etymotic ERA 3A. This insert phone can easily be used together with a cochlear implant (CI) on the contralateral side (and even on the same side). The maximum $2F_0$ distortion component we measured for pure tones of 500 and 1000 Hz at 112 dBSPL was 43 dB below the sound pressure level of the main component.

For electric stimulation we used the Cochlear NICv2 system, which provides a computer interface to several speech processors. The computer was connected to a POD (the clinical fitting device), which was connected to an L34 experimental speech processor. The L34 was programmed to allow streaming arbitrary pulse sequences from the computer. The L34 was connected to the subject's CI via a coil.

The L34 provides a trigger function to synchronize it with other devices. We used the trigger-in function. If this function is enabled, electric stimulation only starts after a trigger signal is received. The second channel of the sound card was used to trigger the L34. The clocks of the sound card and L34 were synchronized by measuring the difference in clock speed and calculating a correction factor to be programmed in the L34.

The subject's own devices (CI speech processor and hearing aid (HA)) were never used because perfect control over the stimulation is required and to avoid device-dependent differences in results.

To control the hardware and conduct experiments, the software platform APEX 3 was developed. Instead of hard-coding all required procedures and stimulation devices, we chose to develop a generic psychophysics platform. While we built on the experience gathered at ExpORL with previous versions of APEX, APEX 3 was completely redesigned and reprogrammed. The remainder of this chapter deals with the development and use of APEX 3.

2.1.2 The software platform: APEX 3

In general, behavioral experiments (e.g. psychophysical experiments or speech perception tests) are controlled by a computer. In most cases custom software is created for each new experiment. However, behavioral experiments have many parts in common. Appropriate stimuli are created and presented to a subject via a transducer, the subject responds via an interface to a computer and the results are stored for analysis. Developing software to perform a specific behavioral experiment is a tedious process that takes a lot of time programming and even more time evaluating all possible response scenarios and eliminating all possible programming errors. Moreover, everything that different experiments have in common has to be programmed and tested again for each different experiment. Consequently, in most cases only researchers with advanced programming skills can set up experiments, whereas there is a strong need for psychophysical testing done by psychoacousticians, audiologists, clinicians, speech scientists, etc., who may have less-advanced programming skills.

A versatile research platform has been developed at ExpORL (Geurts and Wouters, 2000; Laneau et al., 2005) to perform auditory psychophysical and speech perception experiments, either with acoustical stimulation or electrical stimulation via a cochlear implant. Over the years it has evolved from a limited program that could only perform certain specific experiments with electrical stimulation using a cochlear implant of the Laura type (Geurts and Wouters, 2000) to a version that included acoustical stimulation, more extensive procedures and child-adapted test procedures (Laneau et al., 2005), to finally a versatile experimental platform (APEX 3) that allows most auditory behavioral experiments to be performed without any programming, for acoustic stimulation, direct electric stimulation via a cochlear implant or any combination. In this chapter, the novelty of APEX 3 will be discussed. While there are many software packages on the market for visual psychophysics, to our knowledge there are no publicly available packages that are specifically suited for auditory behavioral experiments and that allow many different auditory

experiments to be performed.

The idea behind APEX 3 is that one should be able to set up an experiment quickly without any programming knowledge. APEX 3 is a generic platform with abstract interfaces to the computer monitor, input devices such as keyboard and mouse, and output devices such as sound cards or interfaces to cochlear implants, such that the user can use any of the interfaces without programming any device-specific details. While APEX 3 was mainly developed for research purposes, it is used for rehabilitation and diagnostic purposes too.

APEX 3 is a complete redesign of the previous version of APEX. It builds on the knowledge we gathered during many years of experience with the previous versions of our platform (Geurts and Wouters, 2000; Laneau et al., 2005). The previous versions of APEX have been used in many studies worldwide, as shown by the citations of both APEX papers (Geurts and Wouters, 2000; Laneau et al., 2005). APEX 3 incorporates all features of version 2 (Laneau et al., 2005) and many more. It has already been used at ExpORL for several years and by different international partners. New in APEX 3 is that experiments are now defined in the well-known extensible markup language (XML) format[1], allowing for a structured experiment definition in a generic format. A Matlab toolbox (the APEX Matlab Toolbox (AMT)) is distributed together with APEX 3 to ease the automatic generation of experiment files and analysis of results files. Note that a valid Matlab license is required to use the AMT.

The hardware requirements of APEX 3 are limited to a personal computer running the Linux or Windows operating system and the necessary output devices. The main features of APEX 3 are given in the following list. Features already available in the previous versions of APEX are marked with (*).

- No programming is required to set up an experiment. (*)
- Multiple platforms are supported, including Windows and Linux.
- Multiple output devices are supported, including sound cards, an interface to cochlear implants from Cochlear Corporation and an interface to cochlear implants from Advanced Bionics Corporation. The supported devices can be used in any combination, allowing, for example, for synchronized simultaneous stimulation via a cochlear implant in both ears (bilateral electrical stimulation) or simultaneous stimulation via a cochlear implant in one ear and acoustical stimulation in the other (bimodal stimulation).

[1]The complete XML specification can be found at `http://www.w3.org/TR/xml11/`

- Several psychophysical procedures are readily available and custom procedures can easily be added (plug-in procedure).
- A results file is saved after each experiment. It includes the score, the subject's responses, response times, calibrated parameter values and much more.
- Visual feedback can be given after each trial. (*)
- There is a special animated interface for testing (young) children. (*)
- There is a Matlab toolbox for experiment file creation and advanced result file analysis.
- Custom signal processing filters can be added (plug-in filter).
- Custom interfaces to external controllers can be added (plug-in controller).
- There is a graphical user interface (GUI) for calibration of parameters.

Included with the APEX 3 software package are the following:

- The APEX 3 binaries (the program itself)
- The APEX 3 schema, containing the constraints on the structure of an experiment and documentation for each element
- The AMT, for generating experiment files and analyzing result files
- The APEX 3 user manual
- The APEX 3 reference manual, containing an exhaustive description of all possible elements in an experiment file
- Example experiment files
- Example plug-in procedures, plug-in filters and plug-in controllers

In section 2.2 we describe the general concepts on which APEX 3 is based. In the section 2.3 (design), we show how these concepts are translated to APEX 3 implementation blocks (modules). In section 2.4 the plug-in mechanism is detailed and in section 2.5 it is shown how an experiment can be defined using an XML file. Then, in section 2.6 the general

workflow when deploying APEX 3 is shown and finally, in section 2.7, some examples are given of APEX 3 in use. We will clearly distinguish between the concepts and terminology (section 2.2), and the actual software implementation (the modules, section 2.3).

While a substantial part of our work went into the development APEX 3 and the technical realization of psychophysical tests, it is not necessary to read the current chapter entirely to understand the subsequent chapters.

2.2 Concepts

The design of APEX 3 is based on a few basic concepts that are common to all psychophysical experiments. We define the following concepts: device, controller, datablock, stimulus, filter, screen, response, trial, procedure, experiment, result, ID and parameter.

In section 2.3 we will show how every concept relates to an APEX 3 module.

device is a system connected to the computer that can be controlled by APEX 3. Devices can send signals to a transducer. Examples are sound cards and interfaces to cochlear implants. Devices can have settings (parameters) that can be controlled by APEX 3.

controller is a system connected to the computer that does not accept signals but has parameters that can be controlled by APEX 3. An example is a programmable attenuator that controls the gain of an amplifier.

datablock is an abstraction of a basic block of data that can be processed by APEX 3 and can be sent to the appropriate device. In the case of a sound card, the datablock would be an audio signal in the form of a series of samples that is commonly stored on disk as a so-called wave file.

stimulus is a unit of stimulation that can be presented to the subject and to which the subject has to respond. In the simplest case it consists of a single datablock that can be sent to a single device. More generally it can consist of any combination of datablocks that can be sent to any number of devices, simultaneously or sequentially.

filter is a data processor that runs inside APEX 3 and that accepts a block of data, e.g., a certain number of samples from an audio file, and returns a processed version of the block of data. An example is

an amplifier that multiplies each sample of the given block of data by a certain value.

screen is a GUI that is used by the subject to respond to the stimulus that was presented.

response is the response of the test subject. It could for example be the button that was clicked or the text that was entered via the screen.

trial is a combination of a screen that is shown to the subject, a stimulus that is presented via devices and a response of the subject. Note that while a trial contains a stimulus, it is not the same as a stimulus.

experiment consists of a combination of procedures and the definition of all modules that are necessary to conduct those procedures.

procedure controls the flow of an experiment. The procedure determines the next screen to be shown and the next stimulus to be presented. Generally a procedure will make use of a list of predefined trials. The general working of a procedure is shown in figure 2.2.

result is associated with an experiment and contains information on every trial that occurred.

ID is a name given to a module defined in an experiment. It is unique for an experiment. If, for example, a device is defined, it is given an ID by which it can be referred to elsewhere in the experiment.

parameter is a property of a module (e.g. a device or a filter) that is given an ID. A filter that amplifies a signal could, for example, have a parameter with ID `gain` that is the gain of the amplifier in dB. The value of a parameter can be either a number or text.

Parameter is one of the most important concepts of APEX 3. There are two types of parameters: fixed parameters and variable parameters. A fixed parameter is a property of a stimulus. It cannot be changed by APEX 3 at runtime and is defined when the experiment file is created. It can be used by the procedure to select a stimulus from a list, it can be shown on the screen or it can be used as a piece of information when analyzing results.

A variable parameter is a property of a module of and its value can be changed at runtime. In general, a module can both have variable parameters and set variable parameters of other modules. Examples of modules that can have variable parameters (to be set

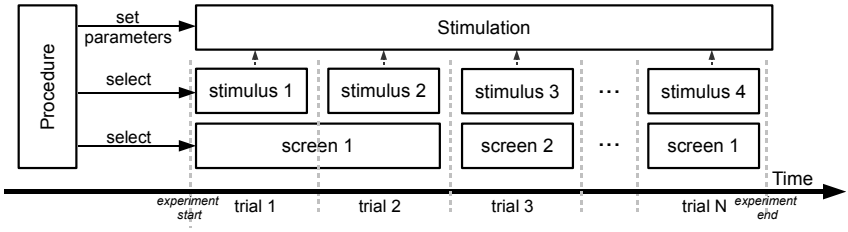


Figure 2.2: Overview of the general working of Procedure. Procedure presents a trial by selecting a Stimulus to be sent to the stimulus output logic and a Screen to be shown.

by another module) are Filter, Controller and Device. Examples of modules that can set variable parameters are AdaptiveProcedure, Device, Calibrator and Screen (more information in section 2.3). If a stimulus description contains a variable parameter, the parameter will be set by Device just before the stimulus is presented.

2.3 Design

Internally, APEX 3 consists of several modules that correspond to the concepts defined in section 2.2. APEX 3 is written entirely in the C++ language[2] and makes extensive use of the Qt library[3]. C++ is an object oriented programming language, and as is usually done in such languages, every module has a base class from which several children (implementations) are derived. For example there is a generic Device module from which the WavDevice module and the L34Device (cochlear implant) module are derived for output via a sound card and output via the Cochlear Corporation nucleus implant communicator (NIC) interface, respectively.

In the following sections a number of modules are described briefly and some of the current implementations are listed. Figure 2.3 gives a graphical overview of some APEX 3 modules. This list of modules is not exhaustive, but is provided to illustrate general principles. Also, since APEX 3 is designed to be easily extended by the developers and third parties (by

[2]The C++ standard is defined in ISO/IEC 14882:1998 and can be found on `http://www.open-std.org/jtc1/sc22/wg21/`

[3]Qt is a programming library created by TrollTech and available from `http://trolltech.com/products/qt/`

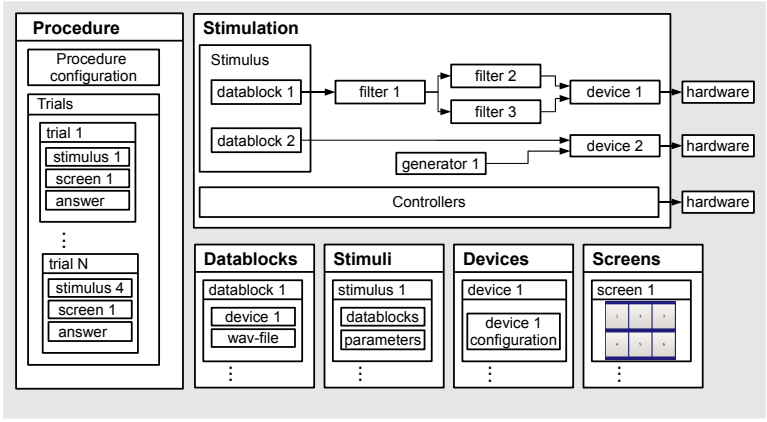


Figure 2.3: Overview of several APEX 3 modules. The stimulation box is not an APEX 3 module, but groups all stimulation-related modules. The four bottom right boxes do not show a complete description of datablocks, stimuli, devices and screens, but serve to guide the eye and indicate that the corresponding modules are defined.

the use of plug-ins), an ever increasing number of modules may be available in the future. The standard set of modules is described fully and exhaustively in the documentation that accompanies the software.

2.3.1 ApexControl

ApexControl is automatically loaded when APEX 3 is started. It takes care of loading all other modules and controlling the general flow of an experiment. ApexControl performs several actions (1) at the start of an experiment, (2) during an experiment and (3) at the end of an experiment. For example it will (1) prompt the user for an experiment to be loaded, (2) ask Procedure to present the next trial and (3) ask ResultSink to save the results.

2.3.2 Procedure

Procedure determines which stimulus is to be played next and which screen is to be shown. The general working of Procedure is illustrated in figure 2.2.

Figure 2.3 shows more details of Procedure. A procedure definition consists of a configuration part and a list of trials. Each trial contains references to a stimulus, a screen and an answer.

Currently, the following implementations of Procedure are present in APEX 3: ConstantProcedure, AdaptiveProcedure, TrainingProcedure, PluginProcedure and MultiProcedure.

To select the next trial, ConstantProcedure selects a trial from the trial list. It can choose a random trial from the trial list every time or present the trials in the order in which they were defined in the trial list. It completes the experiment after every trial has been presented a certain number of times. Technically, ConstantProcedure is the simplest procedure implemented in APEX 3. Typically a percent correct score is calculated from the results, or a psychometric function is fitted to the results.

AdaptiveProcedure is the implementation of an adaptive procedure. It works in the same way as ConstantProcedure, but instead of selecting a random trial it can select a trial or a stimulus based on a parameter that is changed according to the subject's last response. If the response is correct, the task is made more difficult and if the response is incorrect, the task is made easier according to a certain strategy. AdaptiveProcedure can adapt either a variable parameter or a fixed parameter. In the case of a variable parameter, the parameter will be set just before the stimulus is presented (in figure 2.2 this is indicated by the "set parameters" arrow). In the case of a fixed parameter, the stimulus with the fixed parameter closest to the desired value is selected from the user defined list of stimuli. Generally, in psychophysics other types of response strategies using the adaptive procedure exist (Leek, 2001). They can be implemented in APEX 3 using PluginProcedure (see below).

TrainingProcedure does the opposite of ConstantProcedure: it selects the next trial by comparing the subject's last answer to the possible answers defined in the different trials and selecting the one that corresponds. It can, for example, be used to make a training experiment to allow the subject to listen to the stimulus corresponding to each button.

PluginProcedure allows a custom procedure to be defined using ECMAScript. More details are given in section 2.4.

MultiProcedure is not a procedure itself, but it is a wrapper for multiple member procedures of the 4 types above. It allows procedures to be interleaved, either by selecting a random procedure for the next trial or by selecting all member procedures sequentially.

2.3.3 Device

Device can perform the following actions: load a stimulus, set a parameter and start the output. It generally loads data from disk and sends it to a transducer. It can have several parameters that control certain aspects of the device. For example, a sound card can have an output gain parameter.

In figure 2.3 the devices are shown at the right hand side of the *stimulation* box. It is clear that they accept data originating from datablocks or filters and send data to external hardware.

Currently, the following Devices are implemented in APEX 3: WavDevice, L34Device and ClarionDevice.

WavDevice is an interface to sound cards, for acoustical stimulation. Any sound card supported by the operating system can be used. The following sound drivers are supported: Portaudio v19[4], ASIO [5] (Windows only), and Jack[6] (Linux only). The ASIO and Jack drivers allow APEX 3 to be used together with real-time signal processing software on the same sound card.

L34Device is an interface to the NIC interface version 2, provided by Cochlear Corporation, for direct electrical stimulation using a cochlear implant. Via the NIC interface, an L34 or a Freedom processor can be controlled to stream arbitrary pulse sequences to the cochlear implant.

ClarionDevice is an interface to the Bionic Ear Data Collection System (BEDCS) software version 1.16 and higher, provided by Advanced Bionics Corporation. It allows the presentation of arbitrary pulse sequences to the CII or HiRes90K cochlear implants.

2.3.4 Controller

Controllers are used to control devices or software outside APEX 3. They can be considered the same as Devices, with the restriction that they do not load data. Therefore the main properties of controllers are parameters. In figure 2.3, the controllers can be found at the bottom of the stimulation box.

Currently, APEX 3 contains the following controllers: PA5, an interface to the TDT PA5 programmable attenuator[7], Mixer, an interface to the

[4] Portaudio is a free, cross platform, open-source, audio I/O library. `http://www.portaudio.com`

[5] ASIO (Audio Stream Input/Output) is an audio transfer protocol developed by Steinberg Media Technologies GmbH.

[6] JACK is a low-latency audio server, written for POSIX conform operating systems. `http://jackaudio.org/`

[7] `http://www.tdt.com/products/PA5.htm`

sound card mixer provided by the operating system, and PluginController, which allows custom controllers to be implemented by third parties. More information on plug-ins is given in section 2.4.

2.3.5 Screen

The Screen module allows the user to define an arbitrary GUI for subject responses by combining a number of predefined building blocks. The building blocks can be divided into two groups. Elements are the actual controls shown on the screen and Layouts specify the way the elements are arranged on the screen.

The main layout types are GridLayout and ArcLayout. GridLayout arranges elements in a regular grid and ArcLayout arranges elements in a (semi-)circle. ArcLayout can be used for localization experiments, as illustrated in section 2.7.6.

The main Elements are those commonly found in GUIs: Button, Label, Textbox, Spinbox and Picture. A special element is Flash, it allows a FLASH[8] movie or animation to be shown instead of a static image. In this way a test can be adapted to the interest of young children and reinforcement can be given after each trial (Laneau et al., 2005). ParameterLabel and ParameterList can be used to show the current value of a parameter on the screen.

If required, the appearance of all screen elements can be completely customized by the use of style sheets[9]. A style sheet can be specified for the whole of APEX 3, for a certain Screen or per element. Examples of properties that can be changed by the use of style sheets are the color, font or position of an element.

2.3.6 ResultSink

After each trial, ResultSink queries all other modules for information to be stored in a results file. When Procedure has finished, it prompts the subject for a file name and saves the results accordingly. Results are stored in the XML format. While it is very well possible to read and interpret the XML results file, in many cases only a small part of the data presented in this file is required to interpret the results. For example, when evaluating the results of an adaptive procedure, one is primarily interested in

[8] `http://www.macromedia.com/software/flash/about/` Macromedia is currently a division of Adobe Systems Inc.

[9] The specification of CSS (cascading style sheets) and more information can be found at `http://www.w3.org/Style/CSS/`

the staircase and not always in the subject response times. To filter out unwanted information, ResultSink performs an XSL transform[10] on the results to extract the information that is required by the experimenter. The results after XSL transformation can be saved to the results file and can also be shown on screen. Even when performing an XSL transformation, the original XML results file is kept and can be consulted if further information is required.

2.3.7 Calibrator

Calibrator provides a GUI for calibrating parameters and saving and applying calibration results. Commonly a parameter such as output gain is calibrated to achieve the desired stimulation level. Any Stimulus defined in the experiment file can be used as a calibration stimulus.

2.3.8 Filters

Filters are used to process data before sending it to a Device. In figure 2.3 Filters can be found in the stimulation box, in between datablocks and devices. Examples of filters that are currently implemented are Amplifier, for amplifying or attenuating sound data, and PluginFilter, an interface for implementing custom filters. More information on plug-in filters can be found in section 2.4.3.

A special kind of filter is a generator, a filter without input channels. Examples of generators that are currently implemented are SineGenerator, NoiseGenerator and DataLoopGenerator. The first two generate respectively sine waves and white noise. DataLoopGenerator loops a given datablock infinitely.

For each Filter or generator it can be specified whether it should keep on running in between trials (while the user is responding) or not.

2.3.9 Connections

If many Datablocks, Filters and Devices are defined, it may not be straightforward for APEX 3 to know how to connect them. Therefore connections can be defined. Any Datablock can be connected to any Filter or Device and any Filter can be connected to any other Filter or Device. In figure 2.3 the arrows between datablocks, filters, generator and devices

[10] XSL transforms are standardized by the W3C consortium and the specification is available at `http://www.w3.org/TR/xslt`

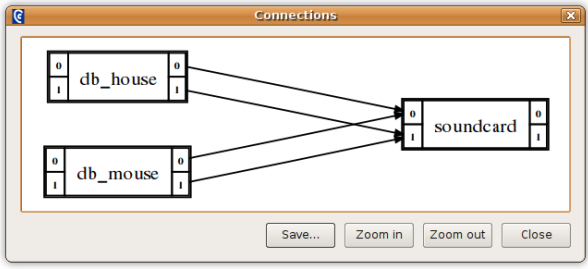


Figure 2.4: Connection graph of the simple example, as generated by APEX 3. In this case each datablock has two channels (left and right) that are connected to the two channels of the sound card. The left and right channels are indicated by the numbers 0 and 1, respectively.

signify connections. By defining connections, a connection graph is created, which can also be shown graphically by APEX 3 for verification purposes. Fig. 2.4 shows the connections for the example experiment of section 2.5.1.

2.4 Extending APEX 3

While APEX 3 can be used for other purposes, it is specifically aimed at auditory research. As research inherently requires "special" and "new" features, it is possible for anyone to extend APEX 3 for their own purposes. Currently APEX 3 can be extended in three different ways, using PluginProcedure, PluginController and PluginFilter.

2.4.1 PluginProcedure

When a plug-in procedure is specified in the experiment file, the user must refer to a script file on disk. In the script file, the user must implement a few functions such as NextTrial, which determines the next screen to be shown and the next stimulus to be played.

The script file is to be written in the ECMAScript language, as defined in the ISO/IEC 16262 standard[11]. ECMAScript was based on the relatively

[11] `http://www.ecma-international.org/publications/standards/Ecma-262.htm`

simple JavaScript language that is used for programming dynamic web pages. Several examples of plug-in procedures are bundled with APEX 3.

While writing such scripts requires some programming, a user need only program the relevant parts of a very specific experiment and not bother with routines that are common to all behavioral experiments, such as output devices, the GUI and saving of results. Programming a simple procedure in ECMAScript typically requires only a few tens of lines of programming code.

2.4.2 PluginController

PluginController allows a user to let APEX 3 control an external device or other software program. As most device manufacturers provide an interface to their devices in the C or C++ language, PluginControllers have to be written in C++. For this purpose the Qt Plug-in mechanism is used and several examples of controllers are provided. Writing a PluginController does not require the user to be familiar with the entire C++ language, but only requires limited knowledge to understand the PluginController examples that are provided and eventual examples from the device manufacturer.

2.4.3 PluginFilter

As the name suggests, a PluginFilter acts like the built-in APEX 3 filters. Just like PluginControllers, PluginFilters have to be written in the C++ language. A PluginFilter is essentially a callback function that is called every time a block of data has to be processed. If implementing a custom algorithm in C or C++ is too bothersome or difficult, a user can alternatively use a different language, such as Matlab or another script language. This option requires that (1) the script language can be called from C or C++, and (2) it is possible to convert between C/C++ data types and the script language's data types.

2.5 Defining an experiment

Previous versions of APEX used a custom text format to define experiments. The format was as simple as possible to enable the creation of experiment files without much technical background knowledge. While APEX 3 of course still has the same aim, it is clear that given the large number of possible experiment configurations, a simple text format does

not suffice. Therefore, the XML format was chosen for defining experiments. To ease the transition, APEX 3 can convert an APEX 2 experiment file to a file in the new XML format.

Advantages of the XML format are that it is human readable, i.e., it can be viewed and interpreted using any text editor, and that it can easily be parsed by existing and freely available parsers[12]. Moreover, many tools exist for editing, transforming or otherwise processing XML files.

Next to adhering to the general XML format, APEX 3 experiment files have a fixed structure that is enforced by an XML Schema[13] file. This file specifies where elements should occur and in addition contains documentation on every element in English. A good XML editor, such as OxygenXML[14] and many others, can use the APEX 3 schema file to check whether an experiment file is valid, to suggest, while typing, what element is to be defined next in the file and to show appropriate documentation per element of the experiment file that is being edited.

In what follows we will describe a very simple APEX 3 experiment file step by step. Note that the order of our descriptions does not correspond to the order of the elements in the experiment file. We will only describe the elements that are necessary to understand the general structure of the file. For more details we refer to the APEX 3 user manual and reference manual, both distributed together with APEX 3. The example is an experiment that will show two buttons on the screen with text "house" and "mouse". When started, it will play either a wave file sounding like "house" or a wave file sounding like "mouse". The subject has to click on the button corresponding to the perceived sound. In speech science, this is called a minimal pair.

An XML file consists of a series of *elements*. Every element can have content. There are two types of content: *simple* content, for example a string or a number, and *complex* content: other elements. An element can also have attributes: extra properties of the element that can be set. Elements are started by their name surrounded by `<` and `>` and ended by their name surrounded by `</` and `>`. In the following example, element `<a>` is started on line 1 and ended on line 7. Element `<a>` contains *complex* content: the elements `<b>` and `<c>`. Element `<b>` contains *simple*

[12] APEX 3 uses the Xerces-c parser for parsing XML files. `http://xerces.apache.org/xerces-c/`

[13] The XML Schema specifications are available at `http://www.w3.org/XML/Schema`

[14] OxygenXML (`http://oxygenxml.com/`) has all necessary features for working with APEX 3 experiment files. It is a commercial program, but a free license can be obtained by non-profit organisations that work in the domains of ecology, human aid and renewable energy sources.

content: the numerical value 1. Element <c> again contains *complex* content: the elements <c1> and <c2>. Element <c1> has an attribute named attrib1 with value 15. Element <c2> on line 5 shows the special syntax for specifying an empty element. This is equivalent to <c2></c2>.

```
1 <a>
2   <b>1</b>
3   <c>
4     <c1 attrib1="15"> </c1>
5     <c2/>
6   </c>
7 </a>
```

As APEX 3 experiment files are in the XML format, the general syntax is the same as in the previous example, but of course the structure is more complex and there are restrictions as to which element can occur where (as enforced by the APEX 3 schema).

2.5.1 A simple example experiment

In what follows we will describe each of the main elements in the experiment XML file separately. Together they define the entire experiment. First we define a device to interface with our sound card.

```
66 <devices>
67   <device id="soundcard"
68     xsi:type="apex:wavDeviceType">
69     <driver>portaudio</driver>
70     <card>default</card>
71     <channels>2</channels>
72     <gain>0</gain>
73     <samplerate>44100</samplerate>
74   </device>
75 </devices>
```

All devices defined in the experiment file are grouped in the element <devices>. As there is only one device in this file, there is only one <device> element. The attribute ID is set to soundcard. As an ID is unique for an entire experiment file, we can use it later on to refer to this device. The xsi:type="apex:wavDeviceType" attribute tells APEX 3 that we are dealing with a sound card. The <device> element contains several other elements that set various parameters of the sound card. The

number of output channels to be used is 2, the output gain is 0 dB and the sample rate is 44100 Hz. Information on all available parameters can be found in the APEX 3 reference manual.

Next we define two datablocks as follows:

```
53 <datablocks>
54   <uri_prefix>../stimuli</uri_prefix>
55   <datablock id="db_house" >
56     <device>soundcard</device>
57     <uri>house.wav</uri>
58   </datablock>
59
60   <datablock id="db_mouse" >
61     <device>soundcard</device>
62     <uri>mouse.wav</uri>
63   </datablock>
64 </datablocks>
```

All Datablock definitions are grouped in the element `<datablocks>`. In this case two datablocks are defined. They each get an ID that is unique for the experiment file and that allows us to refer to them later on. For each datablock, `<device>` refers to the ID of the device that will play the datablock and `<uri>`[15] contains the name of the file from which to read the data. The number of channels in the file is automatically determined by APEX 3. Here we refer to the ID `soundcard` that was defined in the `<devices>` element.

We now have one device with ID `soundcard` and two datablocks with ID `db_house` and `db_mouse`. As no specific connections are defined for this experiment, APEX 3 automatically connects all datablocks to the device. Figure 2.4 shows the connection graph in this case, as generated by APEX 3.

Next we define two stimuli.

```
79 <stimuli>
80   <fixed_parameters/>
81   <stimulus id="stim_house">
82     <datablocks>
83       <datablock id="db_house"/>
84     </datablocks>
85     <variableParameters/>
```

[15] Uniform Resource Identifiers (URI) are defined in RFC 3986. In its simplest form, an URI can be a file name.

```
    <fixedParameters/>
  </stimulus>

  <stimulus id="stim_mouse">
    <datablocks>
      <datablock id="db_mouse"/>
    </datablocks>
    <variableParameters/>
    <fixedParameters/>
  </stimulus>
</stimuli>
```

In this very simple example, each stimulus again gets an ID and refers to one datablock. We now have one device, two datablocks and two stimuli. All stimulation-related specifications are now defined. We proceed by defining a screen.

```
<screens>
  <screen id="screen1">
    <gridLayout height="1" width="2">
      <button row="1" col="1" id="btn_house">
        <text>house</text>
      </button>

      <button row="1" col="2" id="btn_mouse">
        <text>mouse</text>
      </button>
    </gridLayout>

    <buttongroup id="buttongroup">
      <button id="btn_house"/>
      <button id="btn_mouse"/>
    </buttongroup>
    <default_answer_element>
      buttongroup
    </default_answer_element>
  </screen>
</screens>
```

The `<screens>` element can contain several `<screen>` elements. In this case there is only one screen and it contains a GridLayout with a single row and two columns. In the GridLayout, there are two buttons with ID

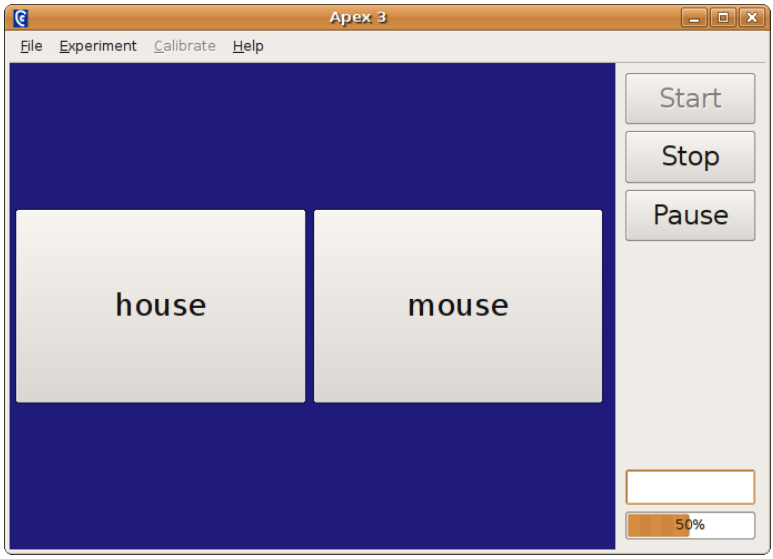


Figure 2.5: Screen of the example experiment

`btn_house` and `btn_mouse`. On each button a piece of text is shown, in this case "house" and "mouse".

The remaining element in `<screen>` groups the buttons into a ButtonGroup. The resulting screen is shown in Figure 2.5. For more information on ButtonGroup we refer to the APEX 3 reference manual.

Finally we define the Procedure that will control the flow of the experiment.

```
7  <procedure
8     xsi:type="apex:constantProcedureType">
9    <parameters>
10     <presentations>2</presentations>
11     <order>sequential</order>
12   </parameters>
13
14  <trials>
15     <trial id="trial1">
16       <answer>btn_house</answer>
17       <screen id="screen1"/>
18       <stimulus id="stim_house"/>
```

```
19      </trial>
20
21     <trial id="trial2">
22        <answer>btn_mouse</answer>
23        <screen id="screen1"/>
24        <stimulus id="stim_mouse"/>
25      </trial>
26    </trials>
27  </procedure>
```

The `<procedure>` element contains two other elements: `<parameters>` and `<trials>` and the attribute `xsi:type="apex:constantProcedureType"` indicates that we use a ConstantProcedure. In `<parameters>` the behavior of the procedure is defined. In this example we specify that each trial has to be presented twice and that the trials are to be presented in the order as specified in the `<trials>` element (sequentially).

The `<trials>` element contains several individual `<trial>` elements that specify a trial. After selecting the next trial to be presented, the Procedure will show the specified screen and send the specified stimulus to the correct devices. After the subject's response, it will check whether the response corresponds to the given answer and decide on the next trial to be presented. For example if the subject clicked on the button with text "house", the procedure will compare the ID of this button (`btn_house`) with the content of `<answer>`.

This simple example illustrates that no programming at all is required to define an experiment and that the syntax is straightforward and easy to learn, especially when using the examples that are provided with APEX 3.

2.5.2 Writing experiment files

For complicated experiments with many stimuli, an experiment file can become rather long and tedious to write manually. There are several solutions to this problem. APEX 3 comes with many examples and most probably one will find an example that can be adjusted to the specific requirements of the experiment. Also, several XML editors can parse the APEX 3 schema file and suggest the element to be defined next and give documentation on the current element in the experiment file.

A more efficient solution is to use the AMT. This toolbox is a collection of Matlab files that generate parts of APEX 3 experiment files. One can use the different functions in the AMT to generate an entire experiment file or one can create a template and fill in the missing parts using the Matlab

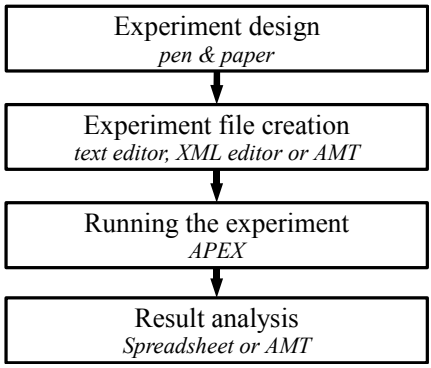


Figure 2.6: Workflow conducting an experiment using APEX 3. AMT is the APEX 3 Matlab Toolbox.

toolbox. Take, for example, the simple experiment from section 2.5.1. If we would like to adapt this experiment to present 50 different words instead of only 2, we could take the original experiment with 2 different words and replace the `<trials>`, `<datablocks>` and `<stimuli>` parts by special markers, e.g., `$$trials$$`, `$$datablocks$$` and `$$stimuli$$`. The AMT contains a function that recognizes these markers and replaces them by given pieces of text. An experiment file with such markers is called a *template*.

Functions like `a3trial`, `a3datablock` and `a3stimulus` in AMT generate the corresponding elements in XML format. We could therefore create a loop in Matlab that is executed 50 times and generates the correct trial, datablock and stimulus elements and afterwards have the AMT replace the markers in our template. A typical Matlab function for generating an experiment file using the latter mechanism requires a few tens of lines of code, in contrast to the thousands of lines of code that would be required to write and debug the same experiment entirely in Matlab.

2.6 Workflow

In this section, we show the typical workflow of setting up, conducting and analyzing an experiment using APEX 3. The workflow is illustrated in figure 2.6.

Experiment design determines the goals and methods of the experiment.

Experiment file creation determines how the methods can be implemented as an APEX 3 experiment by describing them in terms of the basic APEX 3 concepts. If necessary one of many examples can be consulted.

Running the experiment APEX 3 can be used for unattended experiments, where the subjects can respond using a computer mouse, keyboard or touch screen, but also for attended experiments where the experimenter controls the computer. In the latter case, APEX 3 can be configured to show some properties of the current stimulus on screen.

Results analysis For each run of the experiment, a results file is available in XML and, if requested, an XSL transformed version. It is possible to either analyze the results manually by pasting them into a spreadsheet or statistical analysis software, or automatically by using the APEX Matlab Toolbox (AMT) to read the results files and perform advanced analyses.

2.7 Examples

In this section we give a few examples where APEX 3 can be used. This list is nowhere near exhaustive, as APEX 3 is designed to be able to perform any psychophysical experiment.

2.7.1 Gap detection using a 3-alternative forced choice paradigm with a cochlear implant

In our gap detection experiment the method of constant stimuli was used. The subject will, in every trial, hear three different sounds (three so-called intervals). One of the sounds has a small gap in it. The subject has to respond whether the sound with the gap was in the first, second or third interval.

As we want to present the sounds directly to the cochlear implant of our subject, we use the L34Device as a Device to control a cochlear implant from Cochlear Corporation. We need two data files on disk: one containing the sound without gap (NoGap) and one containing the sound with gap (Gap). While our datablocks refer to wave files in the case of a sound card, they now refer to so-called qic files, that can be streamed directly to the cochlear implant and can be created by the Nucleus Matlab Toolbox provided by Cochlear Corporation.

To create the experiment file, we can start from the example in section 2.5.1. First we replace the datablocks by two datablocks that refer to our Gap and NoGap file. Then we replace the stimuli by two stimuli that refer to our Gap and NoGap datablocks and we replace the device by an L34Device. We also change the screen to show three buttons instead of two. Finally we change the procedure to reflect our experimental design. This is done as follows:

```
1  <procedure
2    xsi:type="apex:constantProcedureType">
3    <parameters>
4      <presentations>10</presentations>
5      <skip>0</skip>
6      <order>sequential</order>
7      <choices>3</choices>
8    </parameters>
9
10   <trials>
11     <trial id="trial1" >
12       <screen id="screen1" />
13       <stimulus id="stimulusGap" />
14       <standard id="stimulusNoGap"/>
15     </trial>
16   </trials>
17 </procedure>
```

For experiments where several stimuli are presented during a single trial and the subject is expected to recognize the stimulus that is different in a certain way, multiple stimuli have to be defined per trial. The stimulus that is different is defined using `<stimulus>` and the other stimuli using `<standard>`.

`<choices>` contains the number of stimuli presented to the subject per trial. In this example the number of choices is three, which means that the stimulus defined using `<stimulus>` will be presented once and the stimulus defined using `<standard>` will be presented twice.

Note that while we used the L34Device (not shown in the XML listing) to control the cochlear implant directly, the experiment setup is nearly identical for acoustic stimulation.

2.7.2 Adaptive determination of the speech reception threshold

APEX 3 can be used to determine a subject's speech reception threshold (SRT) for a certain speech material in noise. The SRT is defined as the signal to noise ratio (SNR) at which the subject's performance is 50% correct. We will use an adaptive procedure to determine the SRT. In this example the first speech token (sentence or word) is presented at a low SNR and is repeated at increasingly higher SNRs until the answer is correct. Thereafter the SNR is decreased using a certain step size when the response is correct and increased when the answer is incorrect. Our setup is attended, meaning that the subject has to answer orally and that the experimenter controls the computer running APEX 3. Any speech material can be used. As an example we will use the LIST sentences with the accompanying speech-weighted noise (van Wieringen and Wouters, 2008) which consists of 35 lists of 10 sentences.

Again we start from the example in section 2.5.1. We create a datablock for each of the ten sentences with ID `db-sentenceN` with N the number of the sentence and one extra datablock for the file with speech weighted noise with ID `noisedata`.

We want the noise file to be repeated continuously. Therefore we create a dataloop generator as follows:

```
1 <filter xsi:type="apex:dataloop"
2       id="noisegen">
3   <device>soundcard</device>
4   <channels>1</channels>
5   <continuous>true</continuous>
6   <datablock>noisedata</datablock>
7   <basegain>0</basegain>
8   <gain id="noisegain">0</gain>
9 </filter>
```

The generator has ID `noisegen`, it will use datablock with ID `noisedata` and it will play during the entire experiment, even while the user is responding (line 5). To vary the SNR, in this example we will vary the amplitude of the noise. We will therefore vary gain of our dataloop generator. On line 8 the gain element has an extra ID attribute, which results in the gain of our generator being declared as a parameter that can be modified during the experiment by other APEX 3 modules. In order to change the gain of the dataloop generator, an adaptive procedure is defined. Note that in this case, the level of the noise varies with the SNR

and the level of the speech is held constant. The opposite can be achieved by using an Amplifier to adapt the level of the speech.

```
1  <procedure
2     xsi:type="apex:adaptiveProcedureType">
3  <parameters>
4    <presentations>1</presentations>
5      <skip>0</skip>
6      <order>sequential</order>
7      <nUp>1</nUp>
8      <nDown>1</nDown>
9      <adapt_parameter>
10       noisegain
11     </adapt_parameter>
12     <start_value>-12</start_value>
13     <larger_is_easier>
14       true
15     </larger_is_easier>
16     <repeat_first_until_correct>
17       true
18     </repeat_first_until_correct>
19     <stepsizes>
20       <stepsize begin="0" size="2"/>
21     </stepsizes>
22 </parameters>
23 <trials>
24   <trial id="trial_sentence1">
25     <answer>correct</answer>
26     <screen id="screen"/>
27     <stimulus id="stimulus_sentence1"/>
28   </trial>
29   <trial id="trial_sentence2">
30     <answer>correct</answer>
31     <screen id="screen"/>
32     <stimulus id="stimulus_sentence2"/>
33   </trial>
34 etc...
35 </trials>
36 </procedure>
```

On line 10 the parameter to be adapted is set to the gain of our dataloop

generator by referring to its ID. On lines 7 and 8, the adaptive procedure is defined as a 1up/1down procedure and on line 14 larger values of the parameter are defined to be easier for the subject to respond. The elements `<repeat_first_until_correct>` and `<stepsizes>` on lines 16 to 21 are described in detail in the APEX 3 user manual.

2.7.3 Automatic determination of the SRT

In a clinical setting, the SRT is normally determined with an experimenter (clinician) present. In other situations, such as research or remote tests over the internet, it can be useful or necessary to conduct a test without an experimenter continuously present. To do this with open set materials, i.e., materials for which the subject can respond with any sentence, the subject needs to type the sentence on a keyboard and the computer needs to determine whether the typed sentence was correct. If the subject makes spelling errors, they should not be counted as recognition errors.

In APEX 3 such an automatic open set speech test can be set up by replacing the screen from the previous example by a screen containing a text input field and by replacing the corrector by a block that determines the score, taking into account possible spelling errors. An autocorrection algorithm was developed and is described and evaluated in appendix A.

2.7.4 Evaluation of a signal processing algorithm with an adaptive SRT procedure

Imagine we want to do an SRT test as shown in section 2.7.2, and not only present the stimulus to the subject but first run it through a custom noise suppression signal processing algorithm. In this case we would develop a PluginFilter for our algorithm using the C or C++ language. When a sound signal is played back, APEX 3 splits it in fixed-size blocks of samples and sends each block to the PluginFilter, which can process it. After processing, the resulting blocks are sent to the next Filter or to the output Device.

2.7.5 Bimodal stimulation

In this example, we will use different devices together. We will not create an entire experiment but just create a stimulus that presents an acoustical sinusoid and an electrical pulse train sequentially.

In the `<devices>` element we now have two devices, a WavDevice with ID `soundcard` and an L34Device with ID `l34`:

```
1  <devices>
2    <master>soundcard</master>
3    <device id="soundcard" xsi:type="apex:wavDeviceType">
4      <channels>2</channels>
5      <gain>0</gain>
6      <samplerate>44100</samplerate>
7    </device>
8    <device id="l34" xsi:type="apex:L34DeviceType">
9      <device_id>1</device_id>
10     <implant>cic4</implant>
11     <trigger>in</trigger>
12     <volume>100</volume>
13     <defaultmap> ... </defaultmap>
14   </device>
15 </devices>
```

The `<master>` element indicates that the sound card should be started last. The defaultmap for the L34 is not shown here and for a description of the other L34 parameters we refer to the APEX 3 reference manual.

We create two datablocks: one refers to `sinusoid.wav` and the other to `pulsetrain.qic` and to their corresponding devices. Our stimulus is now defined as follows:

```
1  <stimulus id="stimulus_bimodal">
2    <datablocks>
3      <sequential>
4        <datablock id="db_sinusoid"/>
5        <datablock id="db_pulsetrain"/>
6      </sequential>
7    </datablocks>
8  </stimulus>
```

As the datablocks are inside a `<sequential>` element, first the acoustical sinusoid will be played and immediately thereafter the electrical pulse train will be sent to the subject's cochlear implant. This type of stimulus could for example be used for a pitch matching task or a loudness balancing task with a subject with both an acoustic hearing aid and a cochlear implant. Note that simultaneous bimodal stimulation could be achieved by replacing `<sequential>` by `<simultaneous>` on line 3.

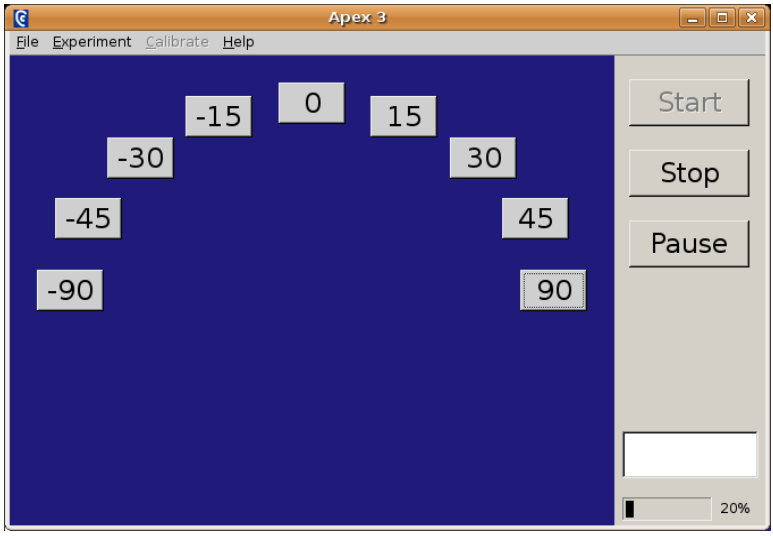


Figure 2.7: Example of an arcLayout with $N = 9$ buttons

2.7.6 Localization of sounds

In a localization experiment, typically the subject is seated in the middle of an arc of N speakers. A stimulus is presented from one of the speakers and the subject's task is to indicate this speaker.

Again starting from the simple example in section 2.5.1, we only need to modify the `<devices>`, `<screens>` and `<connections>` elements.

If the sound card has a sufficient number of output channels to control all the speakers, we only have to change the `<channels>` element in the `<device>` element to value N. If not, multiple sound cards can be used together.

The screen has to be changed to show a semi-circle of N buttons instead of a grid of 2 buttons. Therefore `<gridLayout>` is changed to `<arcLayout>` and the necessary buttons are added. For $N = 9$, the result would look like Fig. 2.7.

2.8 Conclusions

APEX 3 is a versatile program for conducting psychoacoustic behavioral experiments. The most commonly used psychophysical procedures are im-

plemented and APEX 3 can easily be extended with custom procedures. It can control three different output devices: (1) sound cards, (2) streaming and sending pulse sequences to cochlear implants of Cochlear Corporation and (3) sending pulse sequences to cochlear implants of Advanced Bionics Corporation. In addition, custom signal processing algorithms and controllers can be plugged into the APEX 3 framework.

To ease the generation of experiment files and the analysis of results, a Matlab toolbox is provided.

APEX 3 is freely available for anyone after registration. Documentation and many examples are distributed with the software.

Chapter 3

Across-frequency perception of interaural level differences in normal hearing subjects

In current clinical bimodal systems or bilateral cochlear implants (CIs) there is often a mismatch in place of stimulation between the left and right cochlea (see section 1.5). In the current chapter we assess the influence of place mismatch on the sensitivity to interaural level differences (ILDs) in normal hearing (NH) subjects using the test platform described in the previous chapter.

Abstract

Just noticeable differences (JND) in ILD were measured in 12 NH subjects for uncorrelated noise bands with a bandwidth of 1/3 octave and a different center frequency in the two ears. In one ear the center frequency was either 250 Hz, 500 Hz, 1000 Hz or 4000 Hz. In the other ear, a frequency shift of 0, 1/6, 1/3 or 1 octave was introduced. JNDs in ILD for unshifted, uncorrelated noise bands of 1/3 octave width were 2.6, 2.6, 2.5 and 1.4 dB for 250, 500, 1000 and 4000 Hz, respectively. Averaged over all shifts, JNDs decreased significantly with increasing frequency. For the shifted conditions, JNDs increased significantly with increasing shift. Performance on average worsened by 0.5, 0.9 and 1.5 dB for shifts of 1/6, 1/3 and 1 octave. Though performance decreased, the just noticeable ILDs for the shifted conditions were still in a range usable for lateralization. This has implications for signal processing algorithms for bilateral bimodal hearing instruments and the fitting of bilateral cochlear implants.

This chapter is organized in sections introduction (3.1), methods (3.2), results (3.3), discussion (3.4) and conclusions (3.5).

3.1 Introduction

While ILDs for naturally occurring sounds are very small below about 500 Hz, they may be as large as 20 dB at high frequencies (see section 1.6.4 and Moore (2003, Ch.7)). Nevertheless, the human auditory system is able to perceive ILDs in the low frequencies with JNDs as small as ±1 dB, measured with pure tones (Mills, 1960; Yost and Dye, 1988). Low frequency ILD cues are used for localizing nearby sources (Brungart, 1999; Brungart et al., 1999), in the so-called "proximal region", the region within 1 m of the centre of the head.

Mills (1960) presented 5 NH subjects with a reference stimulus with no ILD, followed by a stimulus with an ILD. The stimuli were pure tones. Using the method of constant stimuli, the JND in ILD was determined from half the interquartile separation of the psychometric curves for each subject. JNDs were around 1 dB for 1000 Hz, somewhat smaller for lower frequencies and around 0.5 dB for frequencies higher than 1000 Hz.

Yost and Dye (1988) measured JNDs in ILD for pure tones and different reference signals at 75 % correct, using a linear fit of the psychometric curve. For the reference at ILD = 0 dB they found JNDs of approximately 0.75, 0.85, 1.20, 0.70 and 0.73 dB for 200 Hz, 500 Hz, 1000 Hz, 2000 Hz and 5000 Hz. In the 2AFC procedure, subjects heard one stimulus on the right side and one on the left side and had to respond which one was on the right.

Hartmann and Constan (2002) tested the hypothesis of the level meter model: can the ILD be seen as an integrated measure of stimulus energy, independent of stimulus details? Differences between correlated and uncorrelated stimuli were assessed using white noise and low pass filtered noise (< 1000 Hz). A 2AFC, 1 up/3 down adaptive procedure, targeting the 79 % correct point was used and subjects had to determine the direction of change (right-to-left or left-to-right) for interaurally correlated, anticorrelated or uncorrelated noise. They concluded that the level meter model is sound within half a dB, i.e., the thresholds for each of the tested correlation conditions were within 0.5 dB of each other. JNDs for uncorrelated white noise are in the order of 0.6 dB. For the low pass noise condition they are in the order of 0.9 dB.

In section 1.5.1 some technical problems with bimodal devices are described. The are two main problems for ILD perception. The first problem is that high-frequency ILD cues are absent, because the residual hearing in the acoustically aided ear in many cases does not extend beyond 1000 Hz. The second problem it that there is no established method for matching the place of stimulation between the ears (see section 1.5.2) and the filter

bank used in the CI is not fitted individually. Therefore in most cases there will be a mismatch in place of stimulation between the ears. The same is true for users of bilateral CIs: currently the two CIs are fitted more or less independently and the electrode positions along the left and right basilar membrane are not tuned to the same frequencies.

The aim of this chapter is to assess whether it is possible for NH subjects to perceive ILDs for different degrees of frequency mismatch between the signals in the two ears. Therefore, JNDs in ILD were determined for different frequency shifts in one of the two ears. This was done for different base frequencies, using bilaterally uncorrelated noise band stimuli to simulate the difference in stimulation between the acoustic and electric part of a bimodal system and to eliminate potentially confusing interaural time difference (ITD) cues. Note that uncorrelated stimuli result in a diffuse sound image that is not externalized, i.e., it is perceived inside the head. This makes the task harder (Hartmann and Constan, 2002), but also more realistic when considering binaural bimodal hearing systems, where subjects are presented with largely uncorrelated signals.

Similar work for ITDs was done by Nuetzel and Hafter (1981) and Saberi (1998). They tested subject sensitivity to interaural delay in the envelope of respectively high-frequency amplitude modulated sinusoids and frequency modulated sinusoids and found that as the carrier frequency difference increased, time differences were still detected, but performance dropped rapidly. Given that critical bands in binaural experiments have a bandwidth similar to estimates in monaural experiments (Breebaart et al., 2001; Holube et al., 1998), we expect performance for detecting ILDs to deteriorate when large frequency shifts are introduced.

3.2 Methods

3.2.1 Procedure

General procedure

The JND in ILD was determined for each condition using several runs of an adaptive 1 up/2 down procedure, targeting the 71 % correct point. The procedure determined the ILD of the stimulus that was presented. The start value was 10 dB and the initial step size was 2 dB. After 2 reversals, the step size was decreased to 0.4 dB and after 10 reversals to 0.2 dB. The procedure continued until 12 reversals were obtained. No feedback was given. The mean of the ILDs at the last 6 reversals was taken as the JND

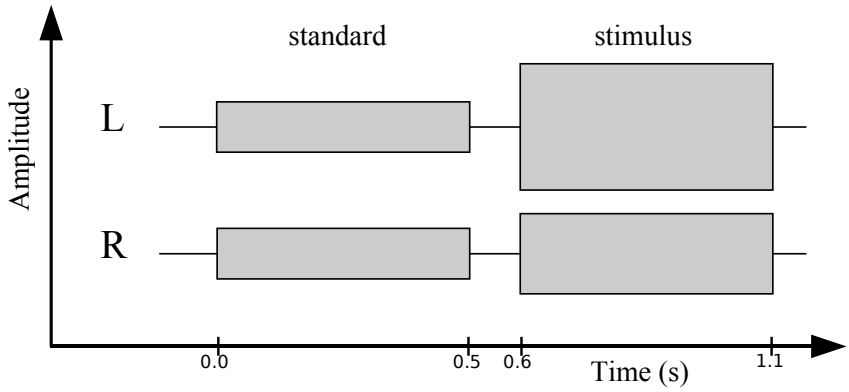


Figure 3.1: Example of a standard-stimulus sequence with a positive rove. For this trial, the correct answer would be "The stimulus sounded on the left hand side of the standard".

for a certain run. If the procedure saturated, i.e., the parameter was 10 dB or 0 dB, the run was discarded and repeated.

In each trial, first a standard was presented, which contained no ILD, followed by a short pause of 0.1 s, followed by the stimulus that contained a certain ILD. The ILD pointed with equal probability to left or right and the magnitude was selected according to the parameter determined by the adaptive procedure. The subjects had to respond whether they heard the stimulus on the left or right side of the standard. One specific case is illustrated in figure 3.1.

Two experiments were done. In the first experiment, to avoid subjects using monaural cues, the overall stimulus level was roved uniformly over ±5 dB. In appendix B we show that in this case, a JND of 4.2 dB could theoretically be attained by only attending to one ear. Because some of the obtained JNDs were larger than 4.2 dB, a second experiment was done with a level rove of ±10 dB.

Subjects were instructed to respond whether they heard the stimulus on the left or right side of the standard. If they were not able to lateralize, they were encouraged to compare the left and right loudness levels. They were also asked to close their eyes during the runs to avoid visual disturbances (there are indications that visual cues can influence responses on localization tests (Lewald and Getzmann, 2006)). They responded using the left and right arrow keys of a computer keyboard. The experiments

were unattended by the experimenter, except for the introduction to the task and regular checks. One run took, depending on the subject, between 78 s and 388 s, with a median of 160 s. This resulted in an average total time of 3.5 h or more of testing per subject in experiment 1, excluding any breaks or short pauses between different runs. The subjects participating in experiment 2, were tested for an additional 1.5 h.

Conditions

JNDs in ILD were determined for 4 base frequencies: 250 Hz, 500 Hz, 1000 Hz and 4000 Hz. The most relevant base frequencies for bimodal hearing are 250 Hz and 500 Hz, because the residual hearing for most subjects that use a bimodal hearing system is restricted to low frequencies. The 1000 Hz and 4000 Hz base frequencies were added as higher frequency reference conditions.

For one adaptive run, the center frequency of the stimulus delivered to ear was always one of the base frequencies and the center frequency delivered to the other ear was the base frequency shifted by 0 oct, 1/6 oct, 1/3 oct or 1 oct. As noise bands of 1/3 oct wide were used, this results in, respectively, full overlap, partial overlap, marginal overlap and no overlap at all of the shifted noise band with the base noise band. The shifts were performed in the upward direction.

Per subject, two base frequencies were selected and all of the shifts were presented for each selected base frequency. A condition consists of a certain base frequency combined with a certain shift. In experiment 1 each condition was presented 8 or 10 times and in experiment 2 it was presented 4 times. To minimize the chance of training effects influencing only a single condition, conditions were always interleaved.

3.2.2 Experimental setup

Stimuli and test setup

The stimuli were 1/3 oct wide noise bands, filtered with a 50th order Butterworth filter to ensure a minimal amount of overlap beyond the cutoff frequencies of the noise bands presented to the two ears. To avoid confusing ITD cues, noise bands were at all time instants uncorrelated between the two ears and new noise bands were generated for each standard and each stimulus. Linear on and off ramping was performed over 0.2 s to avoid clicks and confusing onset cues. The total stimulus duration was 1 s.

For every run, the ear to be presented with the frequency-shifted stimulus was selected at random. On average, each ear was presented an equal number of times with the unshifted stimulus.

To obtain an approximately centered reference signal, the left and right channels were equalized in RMS level with respect to the dBA scale. In this way, the left and right channels sounded approximately equally loud, such that the reference signal was centered in the head. Note that, as a consequence of the dBA weighting, especially at lower frequencies, the levels of the channels differed between ears if measured in dB SPL in conditions with frequency-shifted noise bands.

The ILD was introduced as follows: if S_L and S_R are the levels of the left and right channels of the standard measured in dBSPL, I the ILD to be introduced, r the rove level, randomly selected from the interval $[-5, 5]$ or $[-10, 10]$, and L_L and L_R the levels of the left and right channels of the stimulus, the stimulus was generated according to the following equations (all in dBSPL):

$$L_L = S_L + I/2 + r \tag{3.1}$$
$$L_R = S_R - I/2 + r \tag{3.2}$$

If the same center frequency was presented to both ears, when measuring absolute levels in dBSPL, S_L was the same as S_R and the ILD was I. If different center frequencies were presented, S_L and S_R differed because of the dBA weighting, and the resulting ILD was $I + S_L - S_R$.

All stimuli were presented in a sound booth using the APEX program (see chapter 2 and Laneau et al. (2005)) running on a personal computer, driving a LynxOne sound card that was connected via a mixer to a set of Sennheiser HD250 Linear II headphones.

Calibration of the left and right channels was done by setting the mixer such that a 1/3 oct noise band with a center frequency of 1000 Hz had an overall RMS level of 65 dBSPL. The level of the other stimuli in dBA was equal to the level of the 1000 Hz stimulus in dBA.

Subjects

Twelve subjects participated in experiment 1 and came to the lab for 3 or 4 sessions of 1 to 2 hours. Six of those subjects participated in experiment 2 and came to the lab for an additional 1 or 2 sessions.

All subjects were volunteers and were paid for their cooperation. Their hearing was normal, except for one subject who had a threshold of 40 dBHL

at 4000 Hz. He was only presented with the conditions with base frequencies 250 Hz and 500 Hz and only participated in experiment 1. Two subjects were male and ten were female and all were between 18 and 28 years of age.

3.3 Results

3.3.1 Experiment 1

JNDs in ILD were repeatedly measured for all base frequencies and all frequency shifts. To assess possible training effects, the sequence of results of runs for each frequency/shift condition of each subject is shown in figure 3.2. Each sequence was normalized by dividing by the mean of the last 6 runs in that sequence. The full line connects the averages at each time instant. No clear average long term training effect is evident from this figure. Also, no clear training effect could be seen for any of the subjects separately. As there seems to be a small effect in the first few runs, the first 2 measurements for each condition were discarded from further analysis.

A summary of the results of experiment 1 is presented in figure 3.3. Results are shown for each base frequency and frequency shift, but averaged over all runs and over all subjects. The error bars are at least partly due to inter-subject variance, as opposed to intra-subject variance, as was seen from an ANOVA. The JND in ILD increased with increasing shift (i.e., it is harder to discriminate level differences when the frequencies in the two ears were less similar) and the JND decreased with increasing base frequency (i.e., it was easier to discriminate ILDs when the center frequencies in the two ears were higher). All frequency conditions differed significantly from each other ($F(3, 391) = 25.8, p < 0.00001$ and post hoc tests) as well as all shift conditions ($F(3, 391) = 39.9, p < 0.00001$ and post hoc tests) except for the shifts of 1/3 oct and 1/6 oct.

As the JNDs for the one octave shift conditions are in the neighbourhood of the 4.2 dB value that could theoretically be attained monaurally when using a rove of ±5 dB, this experiment was repeated in experiment 2 with a rove of ±10 dB.

3.3.2 Experiment 2

The small training effect in the first few runs of experiment 1 was not observed in the results from experiment 2. This is probably due to the

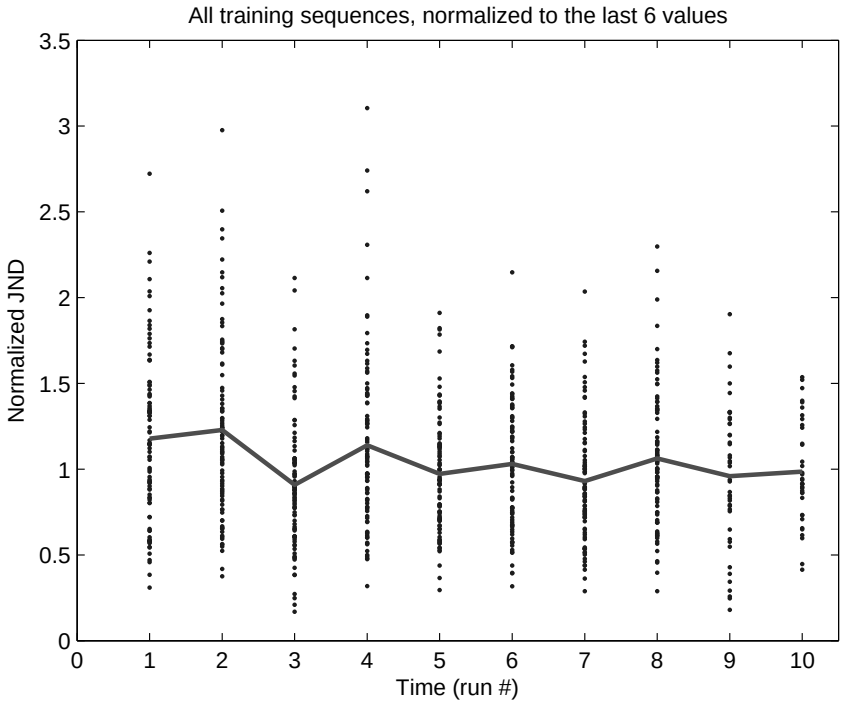


Figure 3.2: All normalized sequences of runs for experiment 1. All values for each sequence were divided by the mean of the last 6 runs for that sequence. Each dot represents the result of an adaptive run. The full line connects the averages at each time instant.

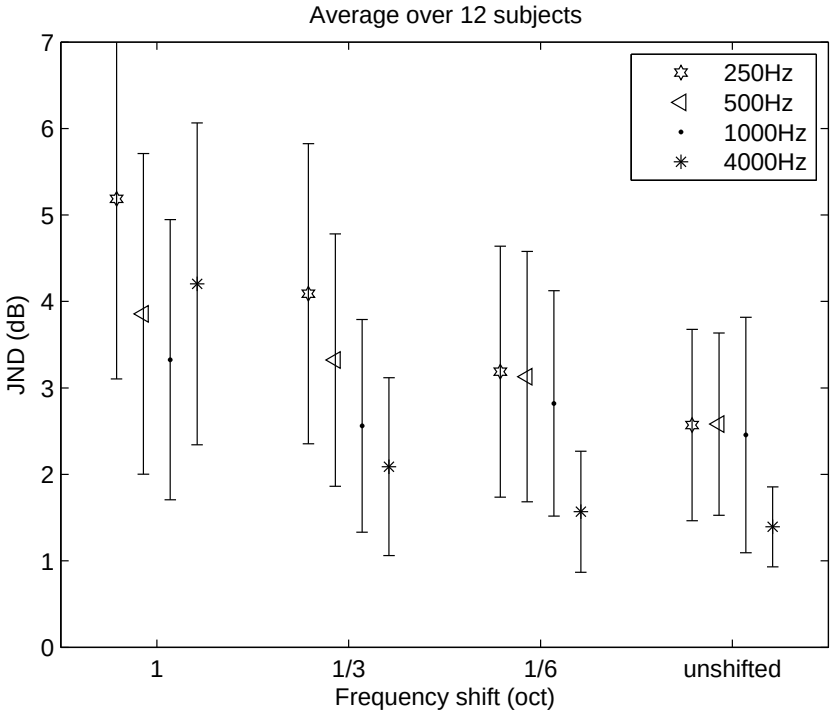


Figure 3.3: JNDs in ILD (in dB) as a function of base frequency and frequency shift for experiment 1 (± 5 dB rove). The total length of the error bar is twice the standard deviation. The data were checked for normality using the Kolmogorov-Smirnov test.

fact that all 6 subjects who participated in experiment 2 also participated in experiment 1 for about 3 hours. Therefore no measurements were discarded from experiment 2 based on training effects.

An ANOVA with factors subject, frequency and shift indicated a significant effect for shift ($F(3, 161) = 24.5, p < 0.0001$). Post hoc analysis with Bonferroni correction showed that all shift conditions differed significantly from each other, except for the shifts of 1/3 oct relative to 0 oct and 1/6 oct relative to 0 oct.

Figure 3.4 shows the differences in threshold values between experiments 1 and 2. On average the JND increased by 0.06 dB from experiment 1 to 2. This difference was, however, not significant in an ANOVA with extra factor experiment. In what follows, we will therefore focus on the results of experiment 1 because it was performed with more subjects and most results are below the 4.2 dB threshold anyway.

3.4 Discussion

Figure 3.3 shows that the JND in ILD increased with increasing shift and the JND decreased with increasing base frequency. The unshifted conditions yielded JNDs of 2.6, 2.6, 2.5 and 1.4 dB for 250, 500, 1000 and 4000 Hz. Hartmann and Constan (2002) reported a JND of 0.6 dB for white noise stimuli and 0.9 dB for low pass noise (< 1000 Hz). Their procedure was similar to ours, but to compare the results, their values have to be multiplied by a factor 2 to compensate for the difference in definition of ILD. Translating their results yields ILDs of, respectively, 1.2 and 1.8 dB. Further differences are due to the fact that, in our experiments, noise bands of a much smaller bandwidth were used. Hartmann and Constan (2002) observed that, for both bandwidths used, JNDs decreased (i.e., performance increased) when the bandwidth increased. Buus (1990) reported that the JNDs for monaural level discrimination decreased when the bandwidth increased. He however used different stimuli: the two ears were stimulated sequentially, while in this study the two ears were stimulated simultaneously.

When considering the results in terms of frequency overlap between the ears, it can be seen that, as soon as the overlap decreased by 1/6 oct, performance decreased significantly. Further decreasing the overlap by 1/3 oct did not yield a significant change compared to the 1/6 oct shift. This can be explained by the fact, that while physically the spectra of the unshifted and 1/3 oct shifted noise band were nearly perfectly separated, there was some spread in the excitation patterns in the cochlea, result-

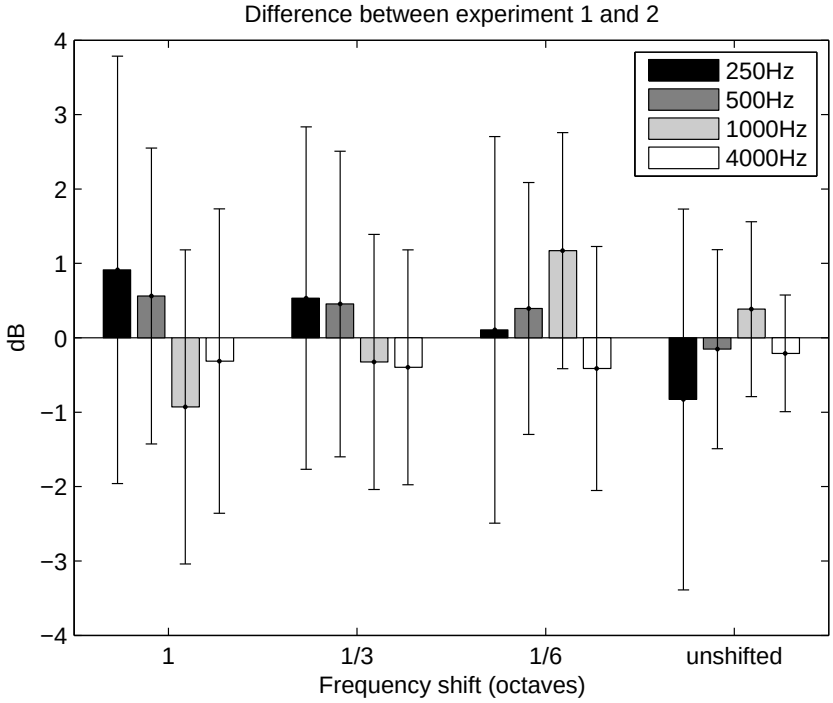


Figure 3.4: Differences between experiment 1 and 2. The bars show the difference in JND. The error bars represent the combined error of both experiments. Positive values indicate that the JND in experiment 1 (±5 dB rove) was larger than the JND in experiment 2 (±10 dB rove).

ing in a certain amount of overlap. The 1 oct shifted noise band yielded significantly worse performance than all other shift conditions, caused by even less overlap in the excitation patterns in the cochlea.

Though significantly larger for the shifted conditions, JNDs were still in a range usable for lateralization of sound sources. The results for the shifted conditions partly confirm the simple level meter model proposed by Hartmann and Constan (2002). The results roughly confirm that the auditory system integrates energy over different frequencies, even over critical band boundaries. However, performance worsened on average by 0.5, 0.9 and 1.5 dB for shifts of, respectively, 1/6, 1/3 and 1 oct, relative to the unshifted condition.

According to Hartmann and Rakerd (1989) the interpretation of our results could be complicated by the fact that the subjects could have ignored the standard that was presented before each stimulus and compared the stimuli to each other, resulting in a larger ILD cue than when comparing the stimulus to the standard. However, this seems unlikely because 1) in contrast to Hartmann and Rakerd (1989), we used level roving, making stimuli with the same ILD sound different, 2) the subjects were repeatedly encouraged to always listen carefully to the standard, 3) an adaptive procedure was used, resulting in a reduction of the effect described by Hartmann and Rakerd (1989) and 4) the results of our unshifted baseline condition correspond well with the results found in the literature. Moreover, even if the absolute values of our results were not accurate, this would not influence the main conclusions, which are based on comparisons between conditions, unless the subjects changed detection strategies between conditions, which seems unlikely.

Though we did not directly measure whether subjects were able to lateralize the stimuli or rather compared level differences between the two ears, we did ask them how they did it for each condition. All 12 subjects reported being able to lateralize in all conditions except for the 1 oct shift. In the 1 oct shift condition, they reported to "sometimes" attend to level differences instead of lateralizing. This attending to level differences can indicate a non-fused image which might be part of the cause of the increased JNDs in the 1 oct shift condition versus the other shift conditions.

3.5 Conclusions

From our JND in ILD experiments with 12 NH subjects, we can conclude that

- ILDs can be detected for uncorrelated narrowband (1/3 oct) noise, with JNDs in the range 1.4 - 5.2 dB
- When a frequency shift is introduced in one ear, ILDs can still be detected, albeit with a slightly higher JND.

The fact that ILDs can be detected across frequencies has important implications for localization using bilateral cochlear implants and contralateral bimodal systems. For bilateral CIs, this means that bilateral matching of electrodes is less important for ILD perception than might be assumed (though performance is still best for the unshifted condition).

For bilateral bimodal systems, this implies that lateralization using ILDs may be improved by introducing or amplifying ILD cues between the acoustical part (the hearing aid) and the low-frequency electrodes of the electrical part. A signal processing system that has access to full band signals of both ears could determine the direction of a prominent sound source and use that direction to calculate a corresponding ILD to introduce at low frequencies. The subject would then have to be trained to localize sound sources using these artificial ILD cues (see chapter 5).

Chapter 4

Perception of interaural level difference and loudness growth with bilateral bimodal stimulation

One of the problems preventing the use of interaural level differences (ILDs) in realistic signals by bimodal listeners is mismatch in place of stimulation in the cochleas (see section 1.5). However, as normal hearing (NH) subjects are sensitive to ILD cues across frequencies (see chapter 3), we can expect bilateral bimodal listeners to be sensitive to ILDs.

Abstract

The sensitivity to ILD in 10 bilateral bimodal subjects was measured. For simultaneous presentation of a pulse train on the cochlear implant (CI) side and a sinusoid on the hearing aid (HA) side, the just noticeable difference (JND) in ILD and loudness growth functions (LGFs) were measured. The mean JND for pitch-matched electric and acoustic stimulation was 1.7 dB. A linear fit of the LGFs on a dB versus μA scale showed that the slope depends on the subjects' dynamic ranges.

This chapter is organized in sections introduction (4.1), methods (4.2), results (4.3), discussion (4.4) and conclusions (4.5).

4.1 Introduction

Two main factors play a role in assessing the utility of ILDs for sound localization in users of a bilateral bimodal system, namely sensitivity to ILD and bilateral loudness growth. By determining the just noticeable difference (JND) in ILD, we can assess whether sensitivity to ILDs is high enough to interpret real-life ILD cues. By assessing bilateral loudness

growth we can assess whether the loudness mapping in current CI speech processors interferes with ILD perception.

Another factor that plays a role is the frequency-to-place mapping. In current CI speech processors, signals are commonly processed in several frequency bands. Each band is then assigned to a certain electrode. In clinical practice the correct tonotopic assignment, which differs across patients, is disregarded (see section 1.5). Therefore, when a narrow band sound is acoustically presented to a bimodal system user, it is likely to be presented to different places in each cochlea. While it has been shown that ILDs can still be detected when such a frequency mismatch is present (see chapter 3 and Francart and Wouters (2007)), it does degrade detection performance and may have an adverse effect on the integration of sounds between ears.

While measures of sensitivity to ILD and interaural time difference (ITD) are not yet available for bimodal listeners, several publications report on localization performance (see section 1.6.2 and Ching et al. (2001); Dunn et al. (2005); Seeber et al. (2004); Tyler et al. (2002)). Because measurement methods differ a lot across studies, it is hard to compare or summarize the results. Overall, most subjects can do side discrimination or lateralization using a bimodal system and only a small fraction of the subjects can do more complex localization tasks. Performance using clinically fitted bimodal systems is generally very limited (see section 1.6.2).

Zeng and Shannon (1992) assessed bimodal loudness growth in three auditory brainstem implant subjects. One subject had normal hearing in the non-implanted ear, while the other subjects had a 40 to 50 dB flat loss at all audiometric frequencies. Loudness growth was measured by sampling equal loudness points between the left and right ear at regular intervals of the total dynamic range. The acoustic stimulus was presented continuously and the electric stimulus was a series of short bursts presented once a second. The subject had to adjust the loudness of the electrical stimulus to the equal loudness point. When plotted on a dB versus μA scale, the LGFs for all three subjects were linear and their slope depended on the dynamic range (DR) of both the acoustical and electrical part.

Eddington et al. (1978) also found a linear dB versus μA relationship with a single subject with a CI. Dorman et al. (1993) came to the same conclusion using one CI subject with a pure tone threshold of 25 dBHL at the test frequency (250 Hz) in the non-implanted ear and a slightly different procedure. The results for CIs therefore seem to correspond to the more extended results for auditory brainstem implants.

Loudness growth using only a CI has been measured by letting subjects estimate the loudness of several stimuli on a scale. Procedures differ be-

tween studies, but commonly the perceived loudness varies exponentially as a function of linear current (Chatterjee et al., 2000; Fu, 2005; Gallego et al., 1999; Zeng and Shannon, 1994).

Reports of JNDs in ILD in NH and bilateral CI subjects are reviewed in section 1.6.5. Summarizing, for NH the JND in ILD is around $1-2\,\mathrm{dB}$ over the entire frequency range. In bilateral CI users performance is worse and varies more across subjects and methods of stimulation. While there have been few studies on lateralization of simple stimuli by hearing impaired subjects (Moore, 1995, p133), performance is not closely related to monaural audiometric thresholds. However, poor performance is usually related to an asymmetric loss.

In this chapter, we assess JNDs in ILD and loudness growth in 10 subjects that used a CI in one ear and were severely hearing impaired in the other ear. First a pitch matching experiment was performed to identify the acoustical sinusoid with the frequency that sounded the most similar to an electrical stimulus presented on the most apical electrode of the CI. Then, loudness balancing experiments were done over the entire acoustic dynamic range. From the found crossover points of the psychometric curves, the LGF can be determined and from the slopes of the psychometric functions the JND in ILD can be found. As a worst case scenario, the experiments were repeated with the most basal electrode. In this way the influence of poorly matched systems (CI and HA) can be assessed.

4.2 Methods

4.2.1 Apparatus

The subject's clinical devices were not used. The test setup consisted of the APEX 3 program and the hardware described in chapter 2. An Etymotic ERA 3A insert phone and an L34 experimental speech processor were used for synchronous stimulation using the CI and the residual hearing. The insert phone was calibrated using a 2cc coupler conforming to the ISO389 standard and the shape of both the electric and acoustic signal were checked using an oscilloscope.

4.2.2 Stimuli

All electrical stimuli were 0.5 s trains of biphasic pulses of 900 pps (pulses per second) with a phase width of 25 µs and an inter phase gap of 8µs. The stimulation mode was monopolar, using both extracochlear reference

electrodes in parallel (MP1+2). These parameters correspond to the clinical maps used by the subjects on a daily basis. The pulse train definitions were generated using custom Matlab scripts and saved to disk. The electrical pulse shapes were generated by the subject's implant and all pulse shape parameters were identical to the settings in the subject's clinical map. We will report electrode numbers in apex-to-base order, such that electrode 1 is the most apical and electrode 22 is the most basal.

All acoustical stimuli were generated using Matlab and were 0.5 s long sinusoids, ramped on and off over 50 ms using a cosine window, to avoid clicks at the beginning and end of the stimulus.

4.2.3 Procedures

Two sets of data were collected. For the first set, the most apical electrode of the CI was used. This electrode stimulates the lowest place-frequency that can be stimulated with the CI. In the second set, the most basal electrode was used that yielded a clear auditory percept and had a minimum dynamic range of 30 CU.

The two electrodes were fitted independently of the clinical fitting. The T (threshold) level was chosen as the just audible level and the C level was the lowest level that was rated as very loud on a 7-interval loudness scale (inaudible - very soft - soft - good - loud - very loud - intolerable). Several parameters for each subject are given in table 4.1.

All procedures were performed for both the most apical (set 1) and most basal (set 2) electrodes. First a pitch matching procedure was done to find the best-matching acoustical pitch for each electrode. Then the frequency of the acoustical stimulus was fixed and several loudness balancing experiments were done to assess loudness growth and JNDs in ILD.

Pitch matching

A pitch matching procedure was used to determine the frequency of the acoustical sinusoid for which the perceived pitch optimally matched the perceived pitch of a pulse train of 900 pps on the selected electrode. At these high rates, the perceived pitch varies only with place and does not depend on variations in the rate (Shannon, 1983; Zeng et al., 2004). Pilot testing with 2 subjects indeed revealed no difference in percept or difference in results from the matching procedure for stimulation at 900 pps or at 7200 pps. Also, as the rate was fixed for all experiments, no influence of rate pitch on our results is to be expected.

					Apical electrode (set 1)				Basal electrode (set 2)			
Subject	Age	M of use	CI side	Etiology	Elec	DR	MF	Thr	Elec	DR	MF	Thr
S1	57	30	R	Progressive	1	45	260	95	22	43	500	85
S2	65	37	R	Noise exposure	1	68	371	95	18	35	626	100
S3	67	46	R	Meniere	1	37	420	100	22	30	250	95
S4	75	33	R	Progressive	1	55	250	90	22	43	250	90
S5	31	28	R	Melas syndrome	1	50	240	95	22	35	350	95
S6	68	89	L	Progressive	1	47	430	75	22	50	870	85
S7	39	59	R	Meningitis	1	68	343	75	22	35	370	85
S8	62	46	R	Congenital	1	47	195	100	22	30	1000	95
S9	31	58	L	Auto-immune	1	70	483	85	20	50	420	90
S10	52	47	R	Congenital	1	47	310	95	22	35	310	95

Table 4.1: Subject information: "Age" is the age in years at the time of testing. "M of use" is the number of months of implant use at the time of testing. "CI side" is left (L) or right (R) (the HA was on the other side). "Elec" is the electrode number (numbered from apex to base) and "DR" is the electrical dynamic range in current units. "MF" is the frequency of the pitch matched sinusoid in Hz and "Thr" is the acoustical threshold in dBSPL

First the acoustical stimuli were balanced in loudness against an electrical stimulus that sounded comfortably loud (the most comfortable level of the electrical stimulus, corresponding to the label "good" on the loudness scale, was determined during the electrical fitting). If the required loudness could not be achieved for some of the acoustical stimuli, the electrical stimulus was reduced in loudness and the balancing procedure started all over again. In this phase, balancing was done by indicating the perceived loudness on the same loudness scale that was used for the electrical fitting. The balancing serves no other purpose than to avoid loudness cues interfering with pitch cues. It has neither relation with nor influence on the loudness balancing experiments performed later in this study.

Second, pitch matching was done using constant stimuli procedures with 4 presentations per stimulus. The electrical and acoustical stimulus were presented sequentially in random order. Every stimulus was presented twice as the first stimulus. The subject had to indicate whether the first or second stimulus sounded higher in pitch. The electrical stimulus was uniformly roved in level over 10 % of the electrical dynamic range, to avoid subjects using residual loudness cues in spite of the loudness balancing previously performed.

A first rough estimation of the matching acoustical pitch was performed by sampling the acoustic frequencies over 2 octaves, spaced by 1/5 oct, ranging from 140 Hz to 560 Hz, resulting in 11 acoustic frequencies. This sampling corresponds approximately to the sampling used by Boex et al. (2006). Then finer scale estimation was performed by using the first obtained rough frequency estimate as the geometrical mean of 11 frequencies over a range of 0.5 oct. The loudness balancing procedure was then repeated for each of these frequencies and the results of both constant stimuli procedures were merged into a single psychometric curve to obtain the best matching frequency. In total, the subject had to answer $11 \times 4 = 44$ times in the rough measurement and another 44 times in the finer scale measurement.

A 2 parameter psychometric function was fitted using a maximum likelihood method to find the 50% point, as well as the slope around the 50% point. An example psychometric function of the fine scale estimation for the most apical electrode of subject S4 is shown in figure 4.1. The first pitch estimate was 280 Hz. Then 11 acoustical frequencies were sampled around this first estimate and a psychometric function was fit to the results. This resulted in a rough estimate of 250 Hz. Then 11 acoustical frequencies were sampled on a finer scale around 250 Hz and again a psychometric function was fitted to the results. The latter function is shown in figure 4.1. The final matched pitch was thus 250 Hz.

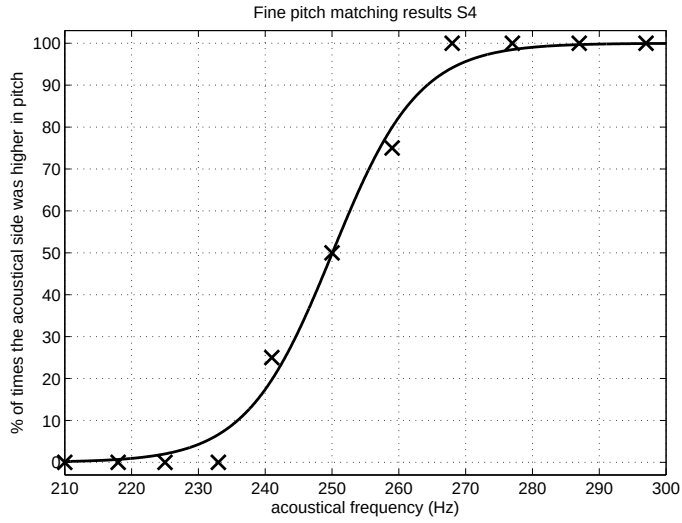


Figure 4.1: Psychometric function for the fine pitch matching experiment for subject S4, set 1.

To confirm the pitch matching, subjects were asked whether the acoustical pitch percept corresponded well to the electrical pitch percept for several intensities. After confirmation, the found pitch was considered correct and used in all subsequent experiments.

Finally the acoustical dynamic range for the matching frequency was determined by finding the acoustical threshold. The upper limit was always the upper limit of the used transducer (112 dBSPL) and was not perceived as uncomfortable by any of the subjects. Note that because of this upper limit of the transducer, the upper limit of the perceptual dynamic range for acoustical stimulation could not be determined.

If no pitch match could be obtained for the most basal electrode, an acoustical frequency was selected for which the dynamic range that could be stimulated was greater than 10 dB and that was subjectively selected as "most similar" to the electrical stimulus. As for most subjects there was no residual hearing at the matching frequency, a lower frequency than the matching frequency was selected. Therefore we can consider set 2 as unmatched, or at least matched worse than set 1.

Loudness growth and JND determination

After the pitch matching procedure, loudness growth and JNDs in ILD were determined by performing several sequential loudness balancing runs. As it was important here to obtain accurate and objective values, a constant stimuli procedure was used instead of the more subjective (but faster) procedure used to balance stimuli before pitch matching.

For set 1 (the most apical electrode), loudness balancing between acoustical and electrical stimuli was done for several electrical levels uniformly spaced over the electrical dynamic range with intervals of 5 % of the dynamic range. In most subjects, for the upper part of the electrical dynamic range, no acoustic amplitude could be found that sounded equally loud (possibly due to the acoustical transducer's upper sound level limit). For the second set (the most basal electrode), loudness balancing was done by sampling the acoustical dynamic range with intervals of at most 5 dB.

If time permitted, more levels were tested, both for set 1 and 2. To avoid subjects being able to answer correctly by using only one ear, different levels for the two ears were presented during the same run. The electrical stimuli were mostly varied in steps of 5 % of the subject's electrical dynamic range and the acoustical stimuli were varied in steps of 2 dB. Step sizes were larger in the first few experiments (to get used to the protocol) and smaller if necessary to find enough points on the slope of the psychometric curve.

In all loudness balancing experiments, the electrical and acoustical stimuli were presented simultaneously, unlike the pitch matching experiments and the LGF experiments for bimodal stimulation found in the literature. The subject was instructed to indicate whether the signal on the left or right hand side was louder. In one run, each stimulus was presented 4 times. When the same stimulus occurred in more than one run, the results of these runs were combined after verifying that they were compatible by overlaying the psychometric curves. There were no disparities within a test session. The subjects performed 2 or 3 sessions of loudness balancing per electrode. For subject S1, the results between the first and second session seemed to differ. Therefore loudness balancing results were only used from the second session with S1 because during the second session much more data were collected than during the first.

To determine the LGF between electrical and acoustical stimulation and the JND in ILD, psychometric curves were fitted for several fixed levels of either the electrical or acoustical part. Psychometric functions were fitted using the psignifit toolbox version 2.5.6 for Matlab[1] which imple-

[1] See `http://bootstrap-software.org/psignifit/`

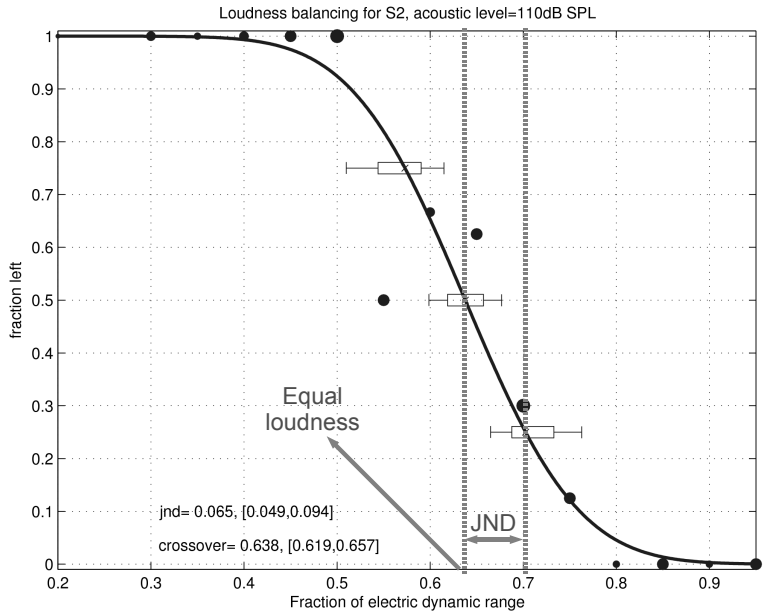


Figure 4.2: Example psychometric function for a loudness balancing experiment for S2, set 2 with the acoustic level fixed at 110 dB SPL. The JND in ILD was 6.5 % of the electric dynamic range and 64 % of the electric dynamic range corresponded to an acoustical intensity of 110 dB SPL.

ments the maximum-likelihood method described by Wichmann and Hill (2001). 68% confidence intervals around the fitted values were found by the BC_A bootstrap method implemented by psignifit, based on 1999 simulations. An example psychometric function is shown in figure 4.2. From the slopes of the psychometric functions, JNDs in ILD were determined as half the difference between the 75% point and the 25% point of the psychometric curve. A 68% confidence interval for the JND was determined by combination of the confidence intervals around these points found by the bootstrap method. To compare JNDs in dB or in percentage electric dynamic range, they can be converted using the found LGF.

For the measurements of set 1 (most apical electrode), the electrical dynamic range was regularly sampled. The corresponding acoustical intensity was determined by fitting a psychometric function for each sampled value of the electrical amplitude.

For set 2 (most basal electrode) on the other hand, the acoustical dynamic range was regularly sampled. Therefore the process was reversed and the electrical value was determined for each sampled value of the acoustical amplitude. All LGFs are shown in figure 4.5 and 4.6.

If for a certain psychometric function the confidence interval could not be determined by the bootstrap method, the point was discarded from all further analyses. This was the case when no data points were available on the slope of the psychometric function (only at the edges). This occurred for only 24 of 186 fits.

The fit of the psychometric function results in several equal-loudness points with error bars. On these points linear regression was performed and R^2 was calculated as an error measure.

4.2.4 Subjects

Ten subjects were recruited amongst the clinical population of the University Hospital of Maastricht (AZM) and the University Hospital of Leuven (UZ Gasthuisberg). All subjects were volunteers and signed an informed consent form. This study was approved by the medical ethical committee. All subjects wore a HA contralaterally to their CI on a daily basis and used a CI of the Nucleus24 type (Cochlear Ltd). S1 and S5 had an electrode array of the Contour Advance type and the other subjects had an array of the Contour type. The clinical processors were of the ESPrit3G type for all but one subject, and of the Freedom type for one subject. All unaided subject audiograms for the acoustically stimulated ear as measured during routine audiometry are shown in figure 4.3. Demographic information for all subjects is given in table 4.1.

The subjects came to the hospital for 4 or 5 sessions of about 2 hours with at least one week and maximally one month between sessions. As the residual hearing of subject S5 abruptly decreased by 10 dB between two sessions, no measurements were made for set 1 for this subject.

Subject S9 had an incomplete electrode array insertion in the cochlea with two electrodes lying outside of the cochlea. All other subjects had normal electrode insertions. Subject S4 had been re-implanted after failure of his first implant, which was implanted in 2002.

4.3 Results

4.3.1 Pitch matching

While for a few subjects some training was needed, pitch matching went smoothly for the experiments with the most apical electrode (set 1). In the next test session, the pitch matching experiment was repeated for verification and the results were always within a few hertz of the previous match. Therefore the results of the first session were used for all subsequent experiments of set 1. The identified frequencies are listed in table 4.1 and shown in figure 4.4.

For the most basal electrode (set 2) however, in many cases no clear pitch match could be found. This is probably due to the lack of acoustical residual hearing at higher frequencies (unaided subject audiograms are shown in figure 4.3). When this was the case, the subject had to select the acoustic frequency that was "most similar" to the electric pulse train. In this case, only acoustic frequencies were presented where the dynamic range was > 10 dB.

For subject S1 the matched pitch for the set 2 experiments was 1124 Hz, but the acoustic dynamic range at this frequency was only 6 dB. We therefore used 500 Hz instead, where the dynamic range was 30 dB. For subject S3 the matched pitch of electrode 1 was 420 Hz. While no good match could be found for electrode 22, 250 Hz was preferred by the subject. Subject S4 reported that the electrical stimulus sounded higher for all acoustical frequencies that could be tested. Therefore 250 Hz was selected in the subsequent tests, based on preference. Subject S7 reported that the electric stimulus was always higher, but preferred the sinusoid of 370 Hz, because it sounded more similar to the electrical stimulus. For subject S9, the matched pitch for the 20th electrode was lower than for the first electrode. This may be due to the subject's partial electrode array insertion, which can cause atypical stimulation patterns when stimulating electrodes

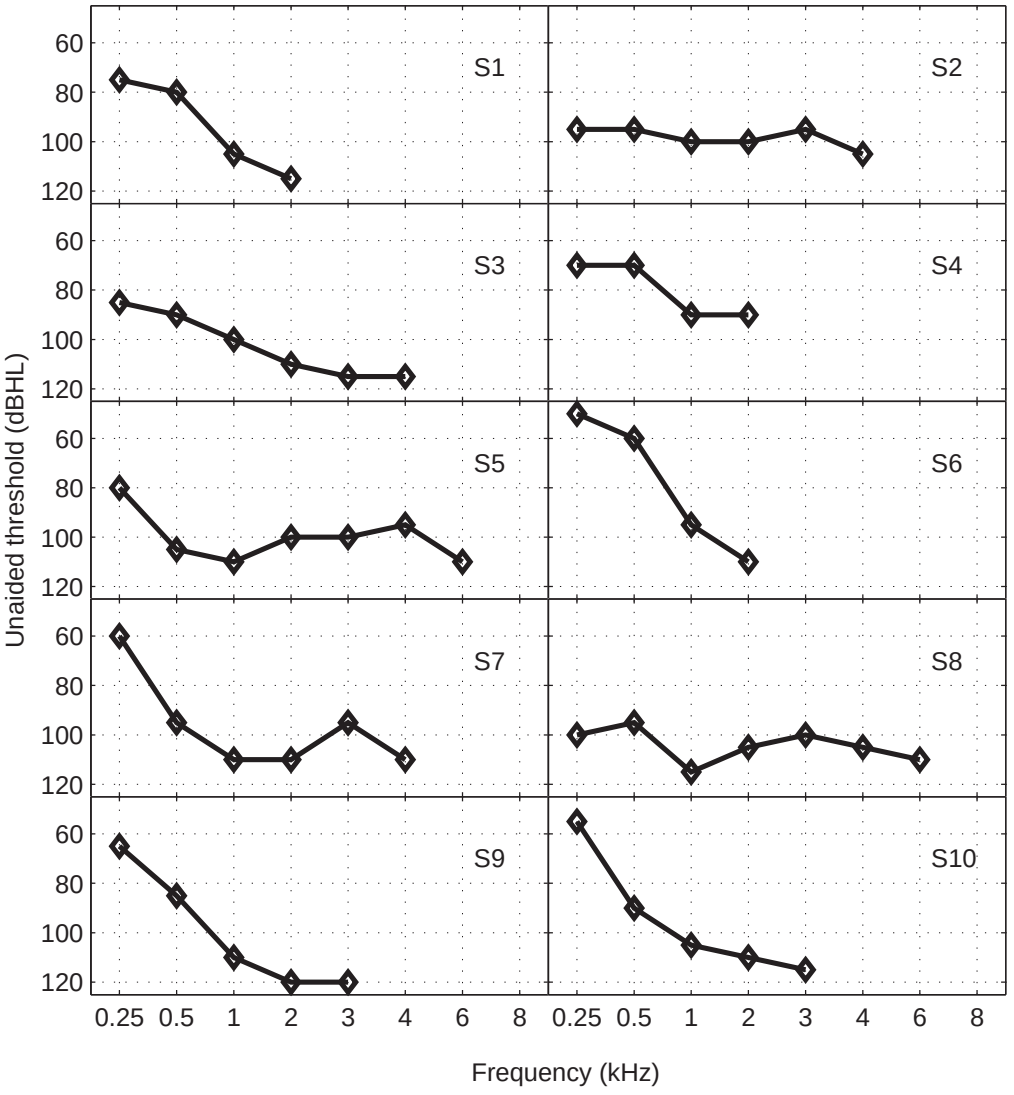


Figure 4.3: Unaided subject audiograms for the acoustically stimulated ear. Note that the vertical axis starts at 50 dBHL. No symbol means no threshold could be found at that frequency.

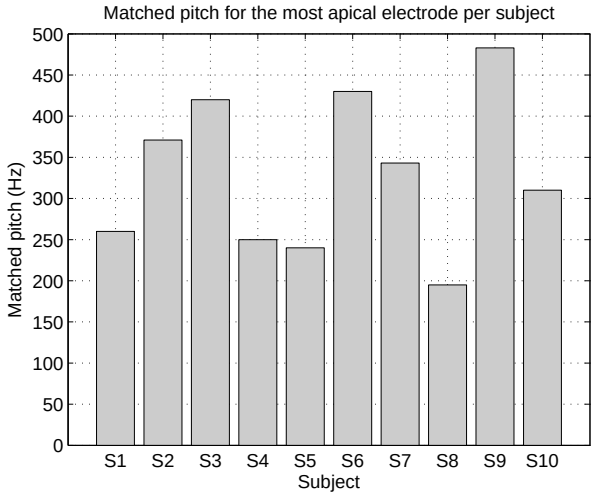


Figure 4.4: Matched pitches for the most apical electrode per subject. Note that S9 has a partial electrode insertion, which explains the higher pitch.

on the edge of the cochlea.

Overall, in set 1, the subjects perceived the acoustical and electrical stimuli as very similar and after some exposure most of them reported the stimuli to fuse to a single percept. One subject (S6) reported that the acoustic stimulus sounded somewhat "warmer". In set 2 however, there was a clear perceptual difference between the stimuli, causing the stimuli not to fuse to a single percept. Therefore, this set should be considered as a worst case scenario, that may occur in practice if the frequency mapping of the CI is very different to the "acoustic" mapping, i.e., when the low frequencies (that can be perceived acoustically) are presented on an electrode that is at a much higher place in the cochlea than the place in the cochlea that is activated by acoustic stimulation.

4.3.2 Loudness growth functions and JNDs in ILD

Based on the loudness balancing experiments, LGFs between electrical and acoustical stimulation were determined. All LGFs are shown in figures 4.5 and 4.6 . The error bars were determined using the bootstrap method. R^2 values are plotted next to each LGF. For each set, a single LGF is plotted

per subject based on linear regression.

From the slopes of the psychometric functions, JNDs in ILD were determined for various intensities. For simplicity we will specify all JNDs in dB change in the acoustical ear for a fixed electrical current in the other ear. Figure 4.7 shows an example set of JND values for the entire dynamic range of subject S6. In this case the median JND was 2.0 dB. Figure 4.8 shows the median of all JNDs for all subjects and both sets. For set 2 the JNDs were converted from current units to dB using the fitted LGF. It can be seen that generally the JND increased when going from set 1 to set 2. The mean JND was 1.7 dB for set 1 and 3.0 dB for set 2. For comparative purposes, some results from the literature on normal hearing subjects and bilateral CIs were also plotted (Laback et al., 2004; Mills, 1960; Senn et al., 2005; Yost and Dye, 1988).

When drawing figures similar to figure 4.7 for all other subjects, it can be seen that for set 1 (apical electrode), the JNDs are very similar over the measured range of intensities. For set 2 however, for some subjects a falling tendency can be observed, i.e., JNDs decrease with increasing sound intensity.

Some subjects found the task subjectively easier for set 2, because they could more easily differentiate between the two ears. However, objectively all of them performed better for set 1. For the latter set, when asked afterwards whether they could hear a single fused sound coming from a certain direction instead of just basing decisions on loudness differences between their ears, 4 subjects answered they could, 2 answered they could not and the 4 other subjects could not answer the question. The fused sound was however not externalized (i.e., was perceived as being located inside the head).

4.4 Discussion

4.4.1 Pitch matching

Boex et al. (2006) reported results of pitch matching of six subjects using Clarion electrode arrays using a similar procedure to the procedure used in this chapter. The matching frequencies for electrode 1 were found to be 460, 100, 290, 260, 288 and 300 Hz. Our results for the most apical electrode (set 1) are in the same range, as can be seen in table 4.1 and figure 4.4. The higher value for subject S9 can be explained by the partial and thus less deep insertion of the electrode array.

For the most basal electrode (set 2) the subjects reported a perceptual

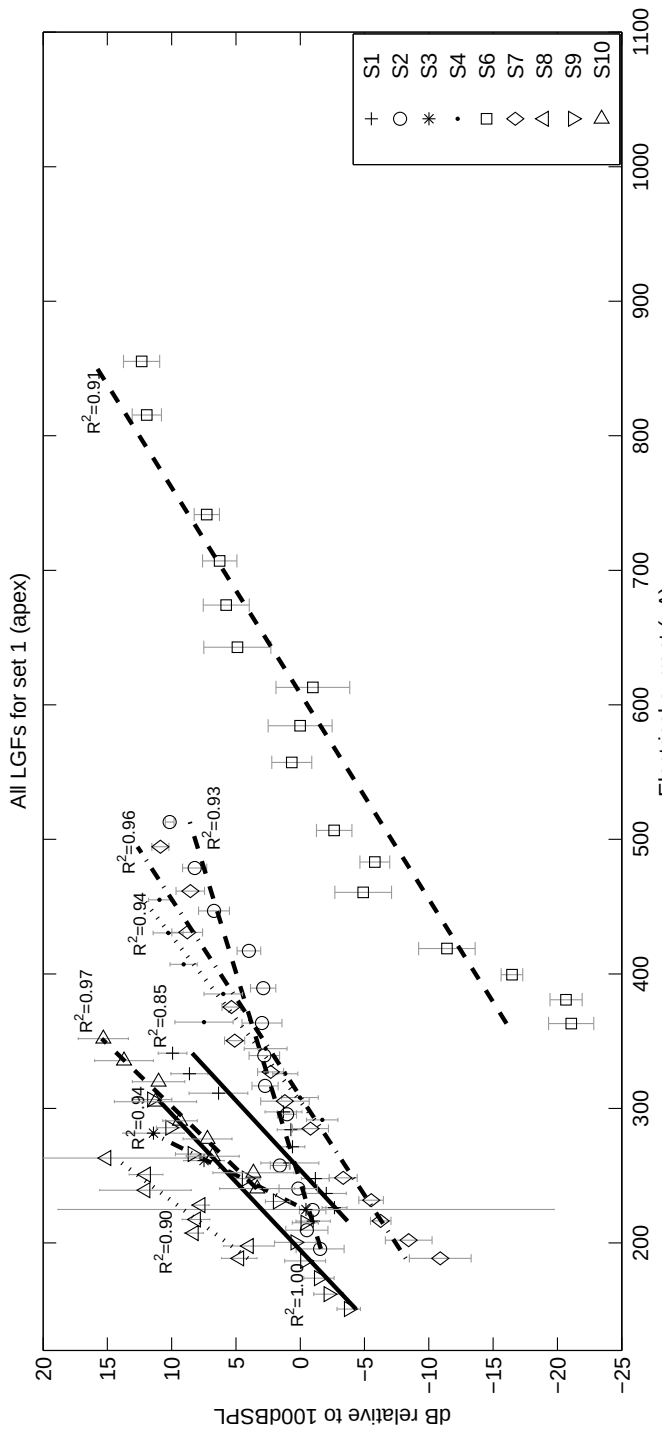


Figure 4.5: LGFs between electrical and acoustical stimulation for set 1. The error bars were determined using the bootstrap method.

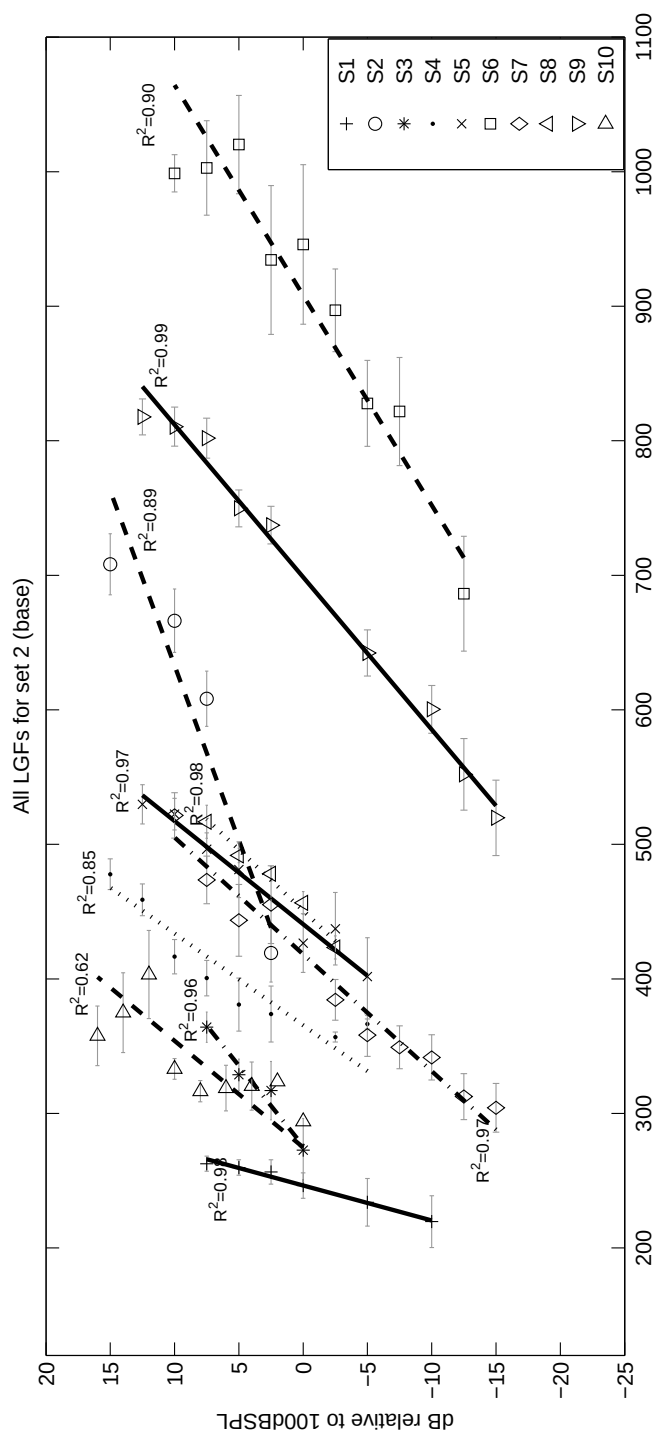


Figure 4.6: LGFs between electrical and acoustical stimulation for set 2. The error bars were determined using the bootstrap method.

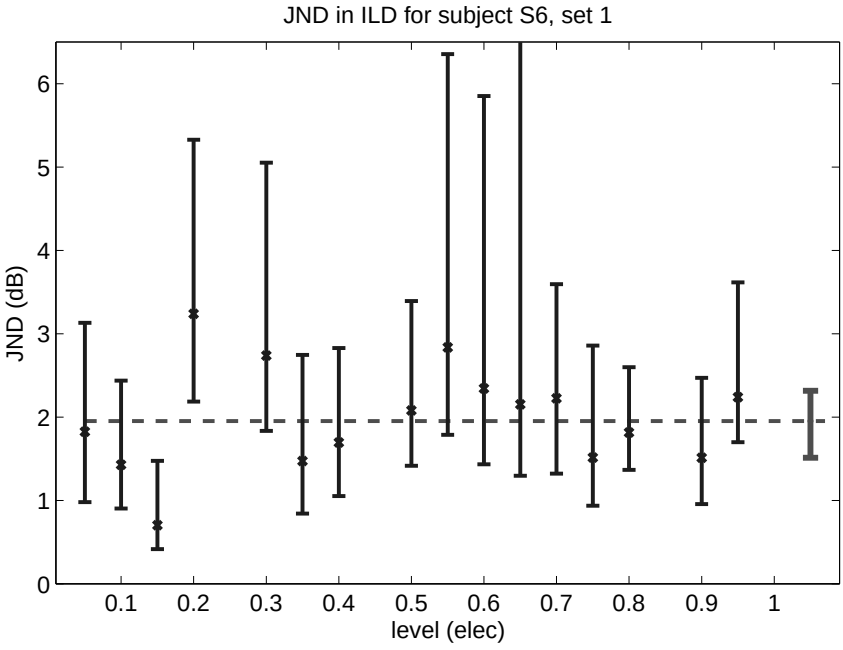


Figure 4.7: JNDs for each electrical intensity for subject S6. The X-axis shows the fraction of the electrical dynamic range in current units. Error bars are 68% confidence intervals determined using a bootstrap method. The dashed line shows the median and the thick error bar on the right hand side shows the 25% and 75% quartiles.

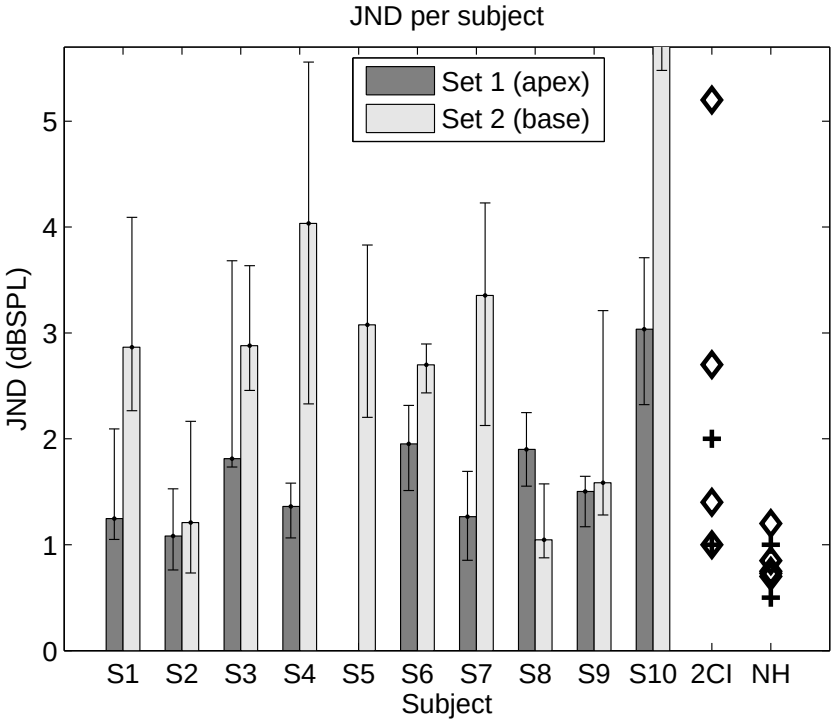


Figure 4.8: All JNDs in ILD expressed as dB change in the acoustical ear for a fixed electrical current in the other ear. The 75% and 25% quantiles are indicated. The JND for S10, set 2 is 7.6 dB. Above the label 2CI, the diamonds show data from Senn et al. (2005) and the plusses show data from Laback et al. (2004) for bilateral CI users. Above the label NH, the diamonds show data from Yost and Dye (1988) and the plusses show data from Mills (1960) for normal hearing listeners.

difference between the stimuli at the two ears and in many cases no clear pitch match could be found. The subjects may have selected the acoustical signal that they could perceive most clearly, i.e., where the dynamic range was sufficient, instead of the signal that was best matched in pitch. Also, a comparison of the obtained frequencies to values in the literature shows that the latter are on average higher or unmeasurable. Boex et al. (2006) reported values of 3050 Hz, 1290 Hz and 1200 Hz for the most basal electrode of Clarion electrode arrays and Dorman et al. (2007b) reported a value of 3400 Hz for the most basal electrode of a MedEl Combi40+ cochlear implant. Therefore the pitches of the electrical and acoustical signal of set 2 of the present study should be considered unmatched. Arguably, the JND results of set 2 would not have been very different if the same acoustical frequency had been used as in set 1.

4.4.2 Loudness growth

Dorman et al. (1993) and Zeng and Shannon (1992) reported a linear growth of the acoustical level in dB versus electrical amplitude in μA. This is not contradicted by our data. A regression line was fitted through all points per subject per electrode and drawn in figure 4.5 and 4.6. R^2 values are shown next to each regression line. However, as the acoustic dynamic range of our subjects was rather small, it is hard to make a strong statement on this topic.

Visual inspection of the set 1 data for subject S6 reveals that an exponential transform of the current may provide a better fit. Indeed, when applying linear regression on the acoustical values in dB versus the electrical values in clinical current units, R^2 increases from 0.91 to 0.95. However, in set 2 for this subject this effect is not observed, neither is there a clear tendency of increase or decrease of R^2 over the other subjects when applying an exponential transform of the current level.

The slopes of the regression lines are subject dependent and depend on both the electrical and acoustical dynamic range (Zeng and Shannon, 1992). In a CI processor the subject's dynamic range is explicitly used when mapping the output of a signal processing channel to an electrode. In current clinical practice it is however not as explicitly used in HA fitting, and the HA is fitted separately from the CI. Therefore a typical fitting of a bimodal system is likely to be suboptimal for ILD perception.

According to Krahe et al. (2000) the use of a ramped acoustical signal and a non-ramped electrical signal could have slightly influenced our results due to possible confounding interaural time difference (ITD) cues. In this case the crossover points we used to determine the LGFs would

all have a slight bias in one direction. We are however confident that our subjects could not perceive ITDs in the present signals because: (1) preliminary tests showed that the results were not influenced by a shift in time of either of the signals; (2) while the electrical and acoustical signal were exactly synchronized in time at transducer level, they were probably not psychoacoustically synchronized. In the acoustical path an extra frequency-dependent delay is present because the sound wave travels through the cochlea while the electrical signal arrives instantly. In the acoustical signal, this extra delay severely degrades ITD cues at the onset of the signals. Moreover, the 50 ms onset ramp in the acoustical signal reduces the salience of possible onset ITD cues. (3) As a high rate pulse train was used electrically and a pure tone acoustically, there were no clear envelope cues in the signals.

4.4.3 Just noticeable differences

The mean JND in ILD of set 1 over all subjects was 1.7 dB. The mean of set 2 was 3.0 dB.

JNDs in ILD for low frequency tones in normal hearing subjects are around 1dB (Mills, 1960; Yost and Dye, 1988). Thus bimodal system users perform slightly worse, but certainly close enough to normal hearing subjects for ILD perception of low-frequency stimuli.

While there was quite some variability amongst subjects, the JND in ILD performance of our bimodal subjects was comparable to the performance of the bilateral CI subjects tested by Senn et al. (2005) and Laback et al. (2004). Here JND values of, respectively, 1.2 dB and $1.4 - 5$ dB were reported.

A comparison of the JNDs of set 1 and set 2 shows an increase in JND for all subjects except S2, S8 and S9. This relates to the fact that for the other subjects the most basal electrode could not as accurately be matched in pitch because of a lack of residual hearing at high frequencies. Francart and Wouters (2007) showed that JNDs in ILD increase in normal hearing listeners with increasing frequency separation between the ears (see chapter 3). This result was confirmed for 2 bilateral CI users by Laback et al. (2004).

4.4.4 Relation to localization performance

While ILD sensitivity is high, performance on localization tasks is still poor for many bimodal system users. This is probably due to three main physical problems with current CI speech processors and HAs.

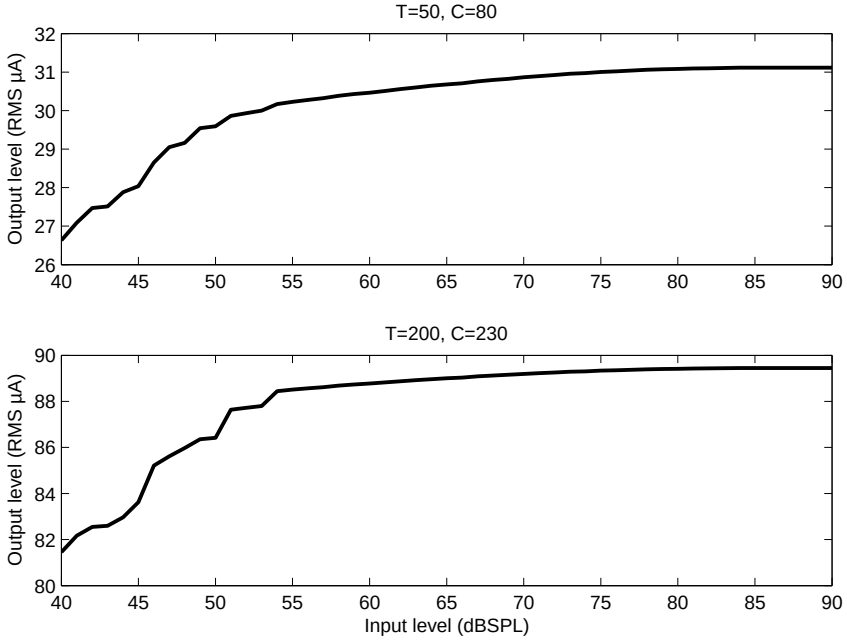


Figure 4.9: Simulated transfer function of the SPrint processor for two different fittings. The abrupt changes in the function are due to quantization into current levels and the breakpoint is due to the saturation level implemented in the loudness growth processing.

A first problem is the absence in real-life signals of large ILDs at low frequencies because the head shadow effect is small for larger wavelengths. While ILDs can reach values of 20 dB at higher frequencies, they are only in the order of a few dB at low frequencies. As most users of a bimodal system do not have residual hearing at high frequencies, they do not have access to clear ILDs. This could be improved by using a signal processing system to amplify ILDs at low frequencies (see chapter 5).

A second problem is the inability to use fine structure ITD cues because they are not transmitted by the CI speech processor and, even if they were, the latency between CI and HA is not optimized.

A third problem is suboptimal bilateral fitting. A CI and HA are in many cases still fitted separately, without paying much attention to loudness balance between both ears. Moreover, compression characteristics are not matched, resulting in unclear ILDs. When considering the transduction of a clinical speech processor, the resulting LGFs (acoustical input versus electrical output) do not have the same shape as the functions found in this paper. Hoth (2007) measured these LGFs between computer controlled stimulation and stimulation via a speech processor with 15 subjects using clinical fitting software for direct electrical stimulation and the subject's well-fitted speech processor and acoustical noise bursts for "acoustical" stimulation. He found that the functions are nonlinear, subject dependent and even electrode dependent within one subject.

To physically assess the loudness transfer characteristics of a clinical speech processor, we made simulations of an example SPrint speech processor using the Nucleus Matlab Toolbox. Figure 4.9 shows the mean output current for a certain acoustical input at the microphone for two hypothetical fittings. It can be seen that the nonlinearity increases when increasing the overall current level. The obtained transfer functions are very similar in shape to the transfer functions found by Hoth (2007). The stimulus was a sinusoid of 250 Hz, chosen to fall exactly in the middle of the first channel of the subsequent processing. The threshold level was set to 50 CU in a first simulation and to 200 CU in a second simulation. The corresponding comfort levels were set to 80 CU and 230 CU. The ACE strategy was used with 8 maxima, the sensitivity was set to 12 and the Q parameter of the loudness growth processing was set to 20. The resulting current level was calculated as the RMS value of all current levels over all channels. Note that this last step implies an over-simplified loudness model, that can however be used as a simple approximation. If the same analysis is done with only one channel selected, it saturates at about 50 dBSPL and does not provide a realistic picture. It is clear that the resulting transfer function is not linear and will most probably inter-

fere with ILD perception in the case of bimodal stimulation. The abrupt changes in the function are due to quantization into current levels, but are on the edge of most subjects' loudness sensitivity, both binaurally and monaurally (Zeng et al., 2004), and thus probably not perceivable. Note that the shape of the transfer function depends on many parameters of the speech processor that can be set in the fitting process. A different combination of T-levels, sensitivity, Q and other parameters will therefore result in either a more linear or an even less linear transfer function.

4.5 Conclusions

LGFs and JNDs in ILD were measured in ten users of a bilateral bimodal hearing system. The LGFs between electric and acoustic hearing can be well approximated by a linear relationship between current in μA and acoustical level in dB. The slope of the line depends on both the electric and acoustic dynamic range and is thus subject dependent. Current CI speech processors use a logarithmic or near logarithmic transfer function whose coefficients depend on various parameters that are set during the fitting to optimize speech perception. This implies that the clinical fitting of the combination of CI and HA will in most cases not be optimal for ILD perception and subsequently binaural lateralization performance.

JNDs in ILD are slightly larger than in normal hearing subjects, but certainly in a range usable for ILD perception. The mean JND for tonotopically matched electrical and acoustical stimulation was 1.7 dB. However, as ILDs are small at low frequencies, for many subjects the use of ILDs will be limited because of a lack of residual hearing at high frequencies. For subjects who do have residual hearing at frequencies where ILDs are present in realistic listening situations, a proper balancing between CI and HA will be important, as the sensitivity to ILDs is high.

Chapter 5

Amplification of interaural level differences for bilateral bimodal stimulation

In chapter 3 it was shown that interaural level differences (ILDs) can be perceived across frequencies in normal hearing (NH) listeners. In chapter 4 it was shown that bimodal listeners are sensitive to ILDs. However, they do not have access to real-world ILD cues because their residual hearing is limited to the low frequencies and ILDs are mainly present at high frequencies (see section 1.5 and 1.6.4). Also, due to several technical problems interaural time difference (ITD) cues cannot be perceived with current clinical bimodal systems (see section 1.5).

Abstract

In this chapter two experiments are described. In the first experiment, headphone simulations of free field listening were used to demonstrate that for normal hearing listeners localization performance based on only ILD cues can under certain circumstances be comparable to their localization performance based on only ITD cues. In the second experiment, using noise band vocoder simulations, it was shown that when using a cochlear implant and a contralateral hearing aid, localization performance can be improved by up to 14° RMS error by artificially amplifying ILD cues at low frequencies. The algorithm that was used for ILD amplification is described.

After the introduction (section 5.1), in section 5.2 the methods common to experiment 1 and 2 are described. In section 5.3 experiment 1 is described and in section 5.4 experiment 2 is described. Finally, in section 5.5 the results of experiment 1 and 2 are discussed.

5.1 Introduction

In chapters 4 and 6 it is shown that users of a bilateral bimodal hearing system are sensitive to the main localization cues, namely ILDs and ITDs. However, current signal processing strategies in cochlear implants (CIs) and hearing aids (HAs) do not allow the use of these cues by the subject. The ITD cues are not available because (1) the CI speech processor removes the fine structure from the signal and (2) the CI and HA are not properly synchronized (see section 1.5.1). The ILD cues are not available because in most subjects residual acoustic hearing is only available at lower frequencies and ILD cues are mainly present at high frequencies (Moore, 2003).

Modification of the CI and HA signal processing to allow ITD perception is still under investigation. Moreover, a minimum level of residual hearing is necessary to allow ITD perception (see chapter 6). However, all subjects in chapter 6 were sensitive to ILDs and while the ILD detection performance decreases with increasing interaural frequency difference, it is even possible to perceive ILDs across frequencies (see chapter 3). Therefore, if clear ILD cues were available between the low frequencies of the residual acoustic hearing and the broader spectrum of the electric stimulation via the CI, bimodal listeners may improve on localization performance.

This chapter assesses via simulations whether localization performance can improve when ILD cues are introduced into the acoustic path. To achieve this, performance of NH listeners on a localization task with only ILD cues is established in experiment 1. In the second experiment, using simulations of bimodal hearing, the performance improvement is measured between conditions with and without application of a practical ILD amplification algorithm.

5.2 General Methods

5.2.1 Simulation of directional hearing

To determine localization performance with manipulated ILD and ITD cues, the method of headphone simulation as described by Wightman and Kistler (1989a) was used. The use of headphone simulations allows the manipulation of ILD and ITD cues in an independent way, which is not possible when using loudspeakers in free field. With the use of head related transfer functions (HRTFs) measured for each subject in the room where the tests take place, localization in the frontal horizontal plane with

headphone simulations is at nearly the same level as with free field stimuli (Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1989b). With the use of non-individualized HRTFs, measured using an artificial head, localization in the virtual field is still possible, but performance decreases (Bronkhorst, 1995; Middlebrooks, 1999; Minnaar et al., 2001; Wenzel et al., 1993). In what follows, stimuli for headphone simulation of free field listening will be called virtual field (VF) stimuli.

To avoid the time consuming process of measuring HRTFs for each subject, HRTFs were measured using an artificial head of type Cortex MK2 and the same set of HRTFs was used for each subject. The use of non-individualized HRTFs measured using an artificial head is known to degrade localization performance (Middlebrooks, 1999; Minnaar et al., 2001; Wenzel et al., 1993), but since we are only interested in differences between conditions and not in absolute performance, this is an acceptable trade-off.

HRTFs were measured for each angle of incidence in an anechoic chamber using exactly the same loudspeaker configuration as in the testing room (see section 5.2.3). The HRTFs were not measured in the testing room because reverberations could result in artifacts in the conditions where ILD and ITD information was removed. Again, the use of HRTFs measured in another room can degrade performance because the visual cues do not match the acoustic cues, but it should not influence the differences between conditions.

Two sets of HRTFs were measured, one set using microphones positioned at the eardrums of the artificial head (in the ear (ITE)) and the other set using omnidirectional microphones positioned on 2 behind the ear (BTE) HA devices which are typically used in high power hearing aids and CI speech processors. The ITE set was used in experiment 1 and the BTE set in experiment 2.

The stimuli were generated by filtering an input signal with the corresponding HRTFs for each angle of incidence. The stimuli were filtered with the inverse transfer function[1], measured between the headphones and the eardrums of the artificial head. This was done to equalize the headphone response and to avoid taking the ear canal into account twice.

5.2.2 Signals

Table 5.1 gives an overview of the signals used in the current study. These signals were used both in free field localization experiments and as in-

[1]The inverse transfer function was determined using an adaptive filter

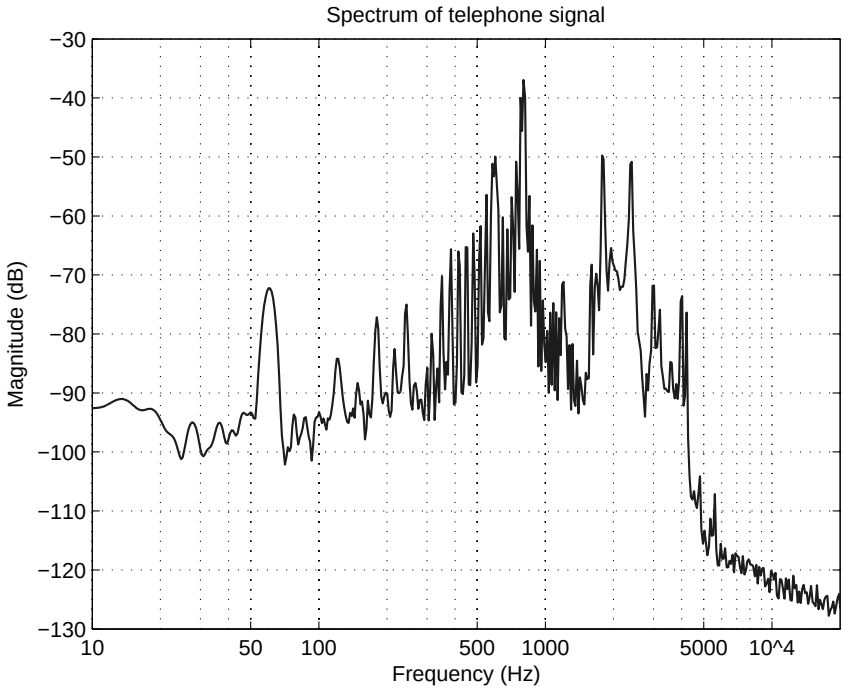


Figure 5.1: Spectrum of the telephone signal

put signals for the virtual field headphone experiments as described in section 5.2.1.

A cosine gate of 50 ms was applied to the start and end of all signals. The telephone signal is the alerting signal of an old-fashioned telephone. Its spectrum is shown in figure 5.1 and its properties are extensively described by Van den Bogaert et al. (2006). An important feature of this signal is the prominent modulation of about 16 Hz.

Note that the signal before any processing will be referred to as *signal* and the signal after processing as presented to the subject will be referred to as *stimulus*. If a signal is presented in free field without further processing, the signal is the same as the stimulus.

Signal	Frequency range	Length
noise14000	$0 - 14000\,\mathrm{Hz}$	400 ms
telephone	mainly $500 - 3000\,\mathrm{Hz}$ (see figure 5.1)	1000 ms
noise3150	1/3 oct around 3150 Hz	400 ms
noise250	1/3 oct around 250 Hz	400 ms
noise500	1/3 oct around 500 Hz	400 ms

Table 5.1: Overview of the signals used.

5.2.3 Apparatus

The subject was seated in a chair in the middle of an array of loudspeakers placed at a distance of 1 m from the subject. The chair was adjusted such that the cones of the loudspeakers were at ear height. Identical loudspeakers were positioned at 15° intervals, yielding a total of 13 loudspeakers, spanning 180° in front of the subject. The loudspeakers were labeled with numbers 1 to 13. In the second half of the circle, the numbers 14 to 24 were attached at 15° intervals. This configuration allows the presentation of stimuli incident from one half of the horizontal plane in free field and responses at locations in the entire horizontal plane (in steps of 15°) in virtual field.

In the free field experiments, active loudspeakers of type Fostex 6301B were used and connected to two 8-channel sound cards of type RME Hammerfall DSP. In the virtual field experiments, headphones of type Sennheiser HD650 were used, connected to one RME sound card.

The experiments were controlled by the APEX 3 program (see chapter 2). The subjects responded using a touch screen and were monitored by the test leader using a microphone and video camera from an adjacent room.

5.2.4 Subjects

Eleven normal hearing subjects aged 21 to 31 years participated in experiment 1 and six participated in experiment 2. Their pure tone thresholds were better than 20 dB HL at the default audiometric frequencies. Experiment 1 consisted of at least two sessions of about 2.5 h and experiment 2 of two sessions of about 1.5 h.

5.3 Experiment 1

Wightman and Kistler (1992) constructed virtual field stimuli with contradicting ITD and ILD cues. They suggested that the ITD is the dominant cue for localization. Macpherson and Middlebrooks (2002) conducted similar experiments but calculated the relative power of ITD and ILD cues to impose bias on lateralization. They concluded that the weight of the ILD is large for a high-pass noise signal, that the weight of the ITD is large for a low-pass noise signal, and that both are important for localizing wide band noise. Experiment 1 replicates part of the latter studies, but adds another condition (VF-ILDonly-RP), in which the ITD is removed by randomizing ITD information across frequencies. This is in contrast with earlier studies, where a bias was imposed by setting the ITD to 0 µs.

5.3.1 Methods

Sound source localization performance was tested for all 5 signals of table 5.1, in 5 different conditions. In the next sections the different conditions (section 5.3.1 and 5.3.1) and the procedures are described.

Conditions and signal processing

In the first condition, the signals were presented in free field (FF), using the 13 loudspeakers. The other 4 conditions (VF-Full, VF-ITDonly, VF-ILDonly and VF-ILDonly-RP) made use of virtual field stimuli, as described in section 5.2.1, plus additional processing as follows.

In the VF-Full condition no further processing was performed.

In the VF-ITDonly condition, ILD cues were removed by setting the magnitudes of the HRTF filters for both ears to one at all frequencies.

In the VF-ILDonly condition, ITD cues were removed by setting them to zero, i.e., by setting the phase of the HRTF filters to zero at all frequencies.

In the VF-ILDonly-RP condition, the ITD zero cue was removed from the stimuli of the VF-ILDonly condition by randomizing the phase across frequencies. This was done using a phase randomization filter. The same filter was used for all stimuli in the VF-ILDonly-RP condition, as described in the next section.

Development of phase randomization filter

In informal pilot experiments, it was observed that if phase differences in a binaural signal are randomized across frequencies, the signal is perceived

as more diffuse, and changes in ITD in the unprocessed signal have no influence on the lateralization of the processed signal. Phase randomization filters were developed as a cascade of 100 digital second order all pass filters for each ear, of the form

$$\frac{r^2 - 2r\cos(\theta)z^{-1} + z^{-2}}{1 - 2r\cos(\theta)z^{-1} + r^2z^{-2}} \tag{5.1}$$

with parameters r and θ. Each of these filters introduces a phase shift of 180° at angle θ with the slope of the phase response related to r. The magnitude response of these filters is perfectly flat. This leaves 200 parameter values to be determined for each cascade of 100 second order filters.

The optimization criterion for the parameters was the amount of variation in ITD over different frequency bands. For a given set of 2 cascades, the left and right ear test signals were filtered by the respective phase randomization filter. The test signal was a white noise signal, filtered by an ITE HRTF recorded at 90°. The result was sent through a gammatone filter bank (Patterson et al., 1995) consisting of 30 filters distributed between 20 Hz and 22 kHz (implemented according to Slaney (1993) with the parameters determined by Glasberg and Moore (1990)). A cross correlation was used to calculate the ITD in each channel. The quality measure was defined as the number of sign reversals of the ITD between the 10 adjacent channels with center frequencies between 100 and 1500 Hz.

A genetic algorithm was used to maximize the quality measure. A random set of parameters was taken, and after calculation of the quality measure, random variations were introduced until the desired value of the quality measure was obtained. This resulted in a set of filters with a quality rating of 9. The corresponding ITDs for each band are shown in figure 5.2. The cross correlation method did clearly not yield meaningful ITDs, as they are very large and are different in nearly every adjacent channel of the gammatone filter bank. Moreover, the maximum of the cross correlation used to determine the ITD was much smaller than in the original signal. The same set of filters was used for each stimulus in the VF-ILDonly-RP condition.

Procedures

Experiment 1 is divided into four parts. Each subject performed the parts in the order specified below. Test and retest were performed on different days. If more than one session was needed for performing all parts once, part 2 was repeated at the start of the next session to ensure that the subject was at the same level of training.

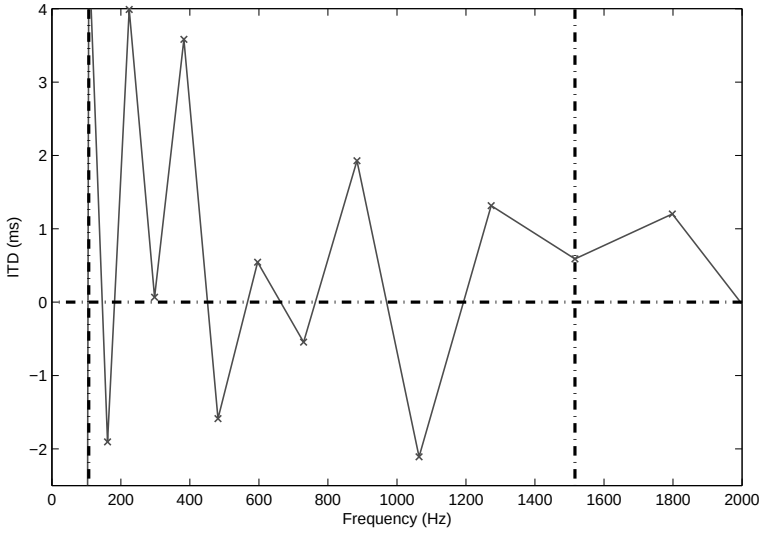


Figure 5.2: ITDs per gammatone filter of the phase randomization filter, determined from the maximum of the cross correlation function between the left and right channel. Every symbol shows the ITD between the corresponding channels of the gammatone filter bank in the left and right ear. The correlation between the two channels was much lower than before application of the phase randomization filter and the cross correlation method did not yield meaningful ITDs.

Parts 1, 2 and 3 served as training or reference conditions for the target experiments of part 4.

1. Assessment of the number of front-back confusions for both free field and virtual field stimuli presented from the right hemisphere (0° to 180°).

2. Familiarization with virtual field stimuli presented from the frontal hemisphere (-90° to +90°).

3. Assessment of localization performance in free field (condition FF) (-90° to +90°).

4. Assessment of localization performance in the different virtual field conditions (VF-Full, VF-ITDonly, VF-ILDonly, VF-ILDonly-RP) (-90° to +90°).

As front-back confusions are common in headphone localization experiments (Wightman and Kistler, 1989b; Zahorik et al., 2006), in part 1 the number of front-back confusions for both free field and virtual field stimuli was assessed. The subject was seated facing the first speaker, and stimuli were presented from the front to the back with 15° intervals in the right lateral half of the horizontal plane. Free field (condition FF) and virtual field (condition VF-Full) stimuli runs were alternated. Only the wide band signals (noise14000 and telephone) were used in this part.

To avoid learning effects during the remainder of this experiment, the subject was further familiarized with the virtual field condition in part 2. The subject was now seated facing the middle of the array (speaker 7) and again only the wide band signals (noise14000 and telephone) were used in the VF-Full condition. The runs were repeated until the RMS error was similar for at least the last 2 runs for both signals.

In the third part, all the signals from table 5.1 were used in free field (condition FF), to establish the baseline localization performance in the test setup.

Finally, in the fourth part, localization experiments were done in virtual field using all signals in all conditions.

In real life, front-back ambiguities are resolved by relying on (1) the shape of the pinnae (Langendijk and Bronkhorst, 2002; Musicant and Butler, 1984; Zahorik et al., 2006), (2) head movements (Bronkhorst, 1995; Wightman and Kistler, 1999) or (3) visual cues. In virtual field, many front-back reversals were expected to occur since (1) the pinnae from an

artificial head were used, (2) subjects were instructed not to move their head (in FF) or head-movement cues were not available (in VF) and (3) no visual cues were given,

To avoid front-back reversals influencing the results, the subjects were given the possibility of responding at angles in the rear hemisphere. Later, front-back reversals were resolved by mirroring answers in the rear hemisphere to the front hemisphere. In what follows, only resolved results will be reported. The chance level, as determined using 10^7 Monte Carlo simulations, was 76.4° RMS error.

To avoid the use of loudspeaker dependent monaural loudness cues, level roving of ±3 dB was used, both in the free field and virtual field conditions. A single run consisted of three presentations of a stimulus from each angle, resulting in a total of 39 presentations per run. For each stimulus presentation a random angle was selected.

Subjects were instructed not to move their head during the presentation of the stimulus. After the presentation they were asked to explicitly look at the apparent stimulus location before responding. This was enforced by monitoring the subjects with a video camera and if necessary asking them to follow the instructions. No feedback was given during the experiment.

The loudspeakers and headphones were calibrated at 65 dB A using a sound level meter of type Brüel&Kjær 2250 and a microphone of type Brüel&Kjær 4192 at the position of the subject's head or an artificial ear of type Brüel&Kjær 4153.

5.3.2 Results

Part 1 – front-back reversals

In part 1 of experiment 1 the percentage of front-back reversals was measured in both the FF and VF-Full conditions. During this part, stimuli were only presented from angles in the right half of the horizontal plane (0° to 180°). The percentage of front-back reversals (P_{FB}) for each subject is shown in figure 5.3. In free field, P_{FB} for the noise14000 signal was very small (median $P_{FB} = 2.6\%$). For the telephone signal P_{FB} was significantly larger ($p < 0.001$, t-test, paired by subject), but still small (median $P_{FB} = 7.7\%$), except for subject N10. As expected, P_{FB} increased in the virtual field condition ($F(1, 75) = 197.0, p < 0.001$, repeated measures ANOVA). It is important to note that no correlation was found between P_{FB} and the localization performance of the different subjects, which suggests that front-back reversals have no influence on the resolved localization performance in the main body of results of experiment 1.

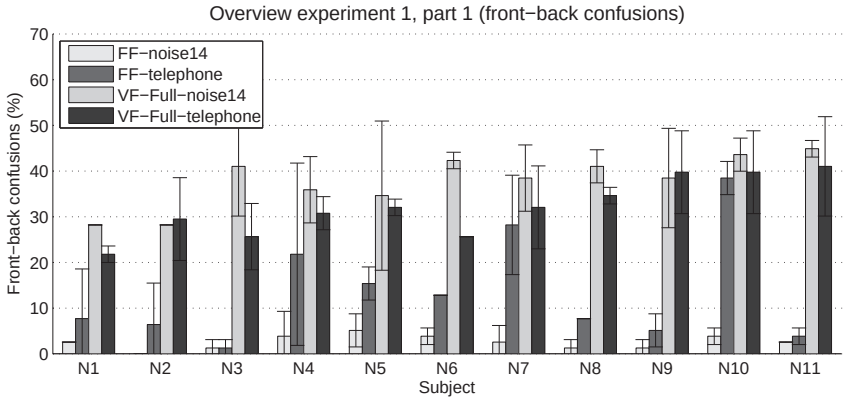


Figure 5.3: Overview results of experiment 1 part 1 – P_{FB} for each of 11 normal hearing subjects. The error bars represent standard deviations on the average of test and retest.

Part 3 and 4 – target conditions

Figure 5.4 shows RMS errors for part 3 and 4 of experiment 1, averaged over all subjects and both runs (test and retest). The error bars represent the between subject standard deviations. Figure 5.5 shows the same results, but now for each angle of incidence of the stimulus. The average bias (not shown) was very small for each combination of stimulus and condition.

An ANOVA with factors subject, condition and stimulus showed a main effect on RMS error of the factors condition ($F(4, 275) = 824.0, p < 0.001$), stimulus ($F(4, 275) = 98.0, p < 0.001$) and subject ($F(10, 275) = 23.7, p < 0.001$). Tukey post hoc tests showed that all individual conditions differed significantly. A second observation was that the stimuli can be divided into 2 groups, the result for which differ significantly: the broadband stimuli (noise14000 and telephone) and the narrow band stimuli (noise250, noise500 and noise3150). Separate ANOVA's and Tukey post hoc tests were carried out for each stimulus separately. In what follows, when a difference is reported as significant, the p-value was smaller than 0.05.

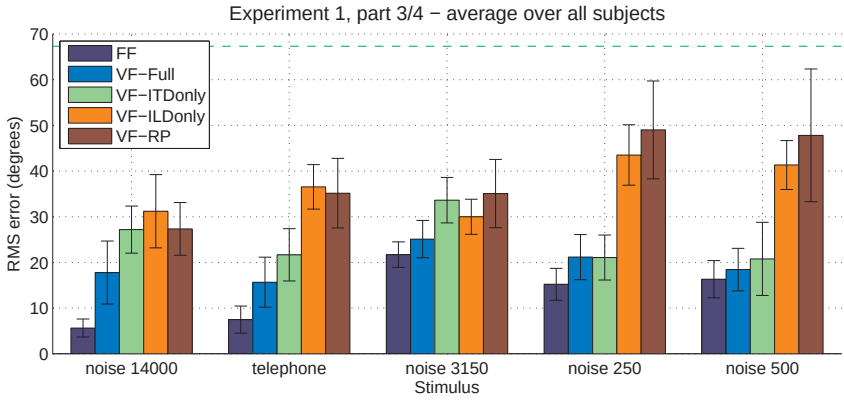


Figure 5.4: Average results for experiment 1 over all subjects (test and retest). The error bars show standard deviations. RMS errors lower than 67.3° are significantly better than chance level (indicated by the dashed line).

5.3.3 Discussion

Free field

The free field results (FF) correspond well with results previously reported for the same test setup (Van den Bogaert et al., 2006). Best performance was achieved when localizing the wide band signals (noise14000 and telephone). Performance was worse for sounds at the sides of the head (larger angles of incidence, see figure 5.5), especially if only ILD cues were available (in the noise3150 signal). This corresponds with the observation that the overall differences in ILD between angles from about 60° to 90° are small, as illustrated by figures 5.6 and 5.9.

Free field versus virtual field

For 7 subjects, the differences between conditions FF and VF-Full were very small for the high-frequency signal (noise3150). Similarly, for 6 subjects, the differences between conditions FF and VF-Full were very small for the low-frequency signals (noise250 and noise500). Other subjects showed larger differences between conditions FF and VF-Full. These differences, which are clearly subject dependent, are probably due to the use of an artificial head instead of individualized HRTFs, which is known to

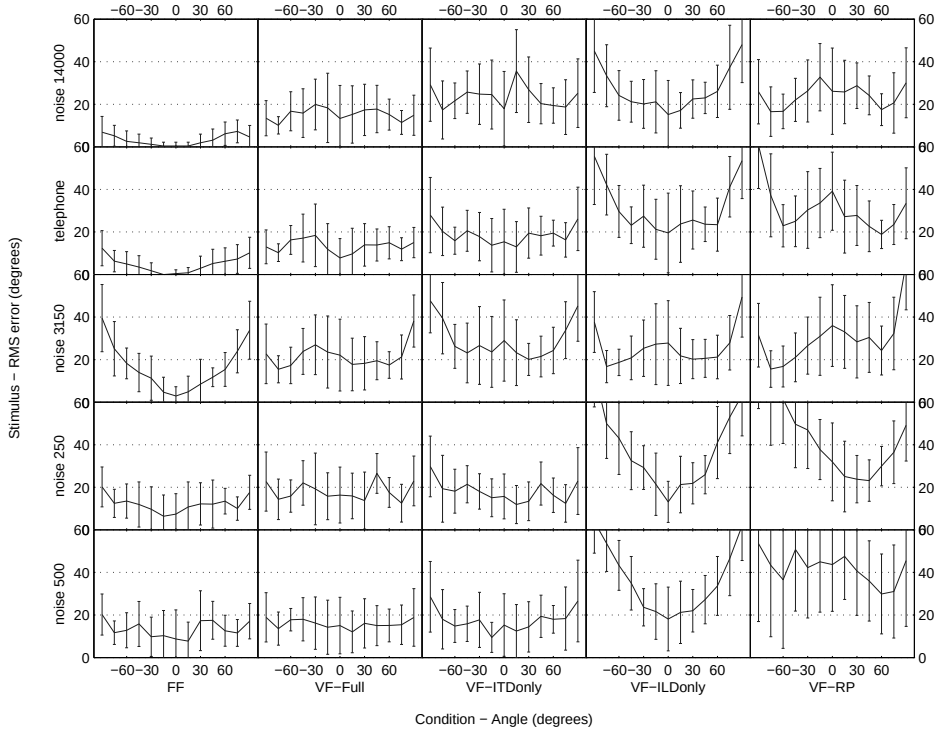


Figure 5.5: Results for Experiment 1 for each angle, averaged over all subjects (test and retest). The error bars are between-subject standard deviations.

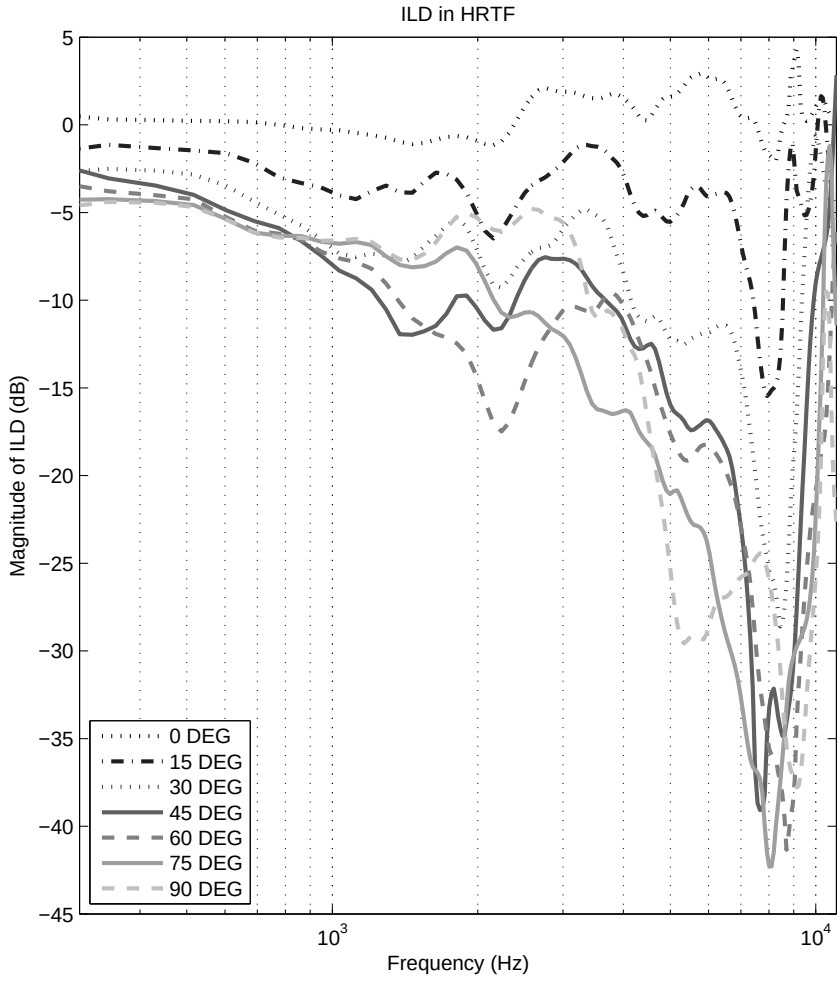


Figure 5.6: ILD for each frequency and angle of incidence, determined from ITE HRTFs, measured using an artificial head.

generate subject dependent differences (Bronkhorst, 1995; Middlebrooks, 1999; Minnaar et al., 2001; Wenzel et al., 1993).

The differences between free field and virtual field varied over subjects, but some subjects obtained very similar scores, both for low frequency and high frequency signals. Therefore the used headphone simulations can be considered valid, especially for comparison between virtual field conditions. In what follows, only differences between virtual field conditions will be considered.

Condition VF-ITDonly

Comparison of conditions VF-Full and VF-ITDonly showed that for the low-frequency signals (noise250 and noise500) performance did not significantly change when removing ILD information. This is due to the fact that at these frequencies localization is dominated by ITD cues (Moore, 2003).

For the signals containing high frequencies (noise14000, telephone and noise3150), performance decreased significantly when removing ILD information. Interestingly, while the noise3150 signal contained only high frequencies, localization was still far from chance level. As fine structure ITD cues are not available at frequencies above 1500 Hz (Moore, 2003), this is probably due to envelope ITD cues. The same is true for the telephone signal which was localized relatively well in condition VF-ITDonly. It does have some low frequency content (see figure 5.1), but more importantly it has prominent amplitude modulations, which are known to increase the salience of the ITD cues (Macpherson and Middlebrooks, 2002).

Conditions VF-ILDonly and VF-ILDonly-RP

For the wide band signals (noise14000 and telephone), performance in the VF-ILDonly condition was worse than in the conditions VF-ITDonly and VF-Full. However, when the ITD zero cue was removed by randomizing the phase (condition VF-ILDonly-RP), performance for the noise14000 signal reached a similar level as for the VF-ITDonly condition (non-significant difference between VF-ITDonly and VF-ILDonly-RP, $p = 0.99$). For the telephone stimulus, the situation is different because the envelope cues in the modulations of the signal were not sufficiently eliminated by the phase randomization filter.

For the high-frequency signal (noise3150), the result for the VF-ILDonly condition was nearly at the same level as for the VF-Full condition and there was a non-significant ($p = 0.054$) trend of improvement compared

to the VF-ITDonly condition. The remaining difference between conditions VF-Full and VF-ILDonly can be attributed to the envelope ITD cue that was available in the VF-Full signal. When comparing conditions VF-ILDonly-RP with VF-ILDonly, it was observed that performance in condition VF-ILDonly-RP was slightly worse. This is probably due to the more diffuse nature of the stimulus in condition VF-ILDonly-RP. Interestingly, the differences between conditions VF-ILDonly-RP and VF-ILDonly were mainly observed at angles around 0° (see figure 5.5).

For the low-frequency signals (noise250 and noise500), performance in both the VF-ILDonly and VF-ILDonly-RP conditions was significantly poorer than in the VF-Full condition since ILDs were very small for low frequencies and the dominant ITD cues (Macpherson and Middlebrooks, 2002) were only available in the VF-Full condition.

The effect of the ITD zero cue generating a bias towards 0° in the VF-ILDonly condition is illustrated by comparison of the results for each angle between conditions VF-ILDonly and VF-ILDonly-RP. For the wide band signals, in condition VF-ILDonly the errors made for stimuli at the sides of the head (from larger angles) were larger than in condition VF-ILDonly-RP, indicating a bias towards 0° for the former.

In the VF-ILDonly-RP condition there was a tendency of increased error at small angles when compared to the VF-ILDonly condition. This was also observed in the previous section and it can probably be attributed to the diffuse nature of the stimuli in the VF-ILDonly-RP condition.

The main result for the current study, however, is that localization is possible with only ILD cues. If the stimulus contains enough high frequencies and no conflicting ITD cues are present, ILD cues can be as useful for localization as ITD cues.

5.4 Experiment 2

In experiment 2 we investigated how localization can be improved for bilateral bimodal stimulation using amplified ILD cues.

5.4.1 Methods

Simulations of bimodal hearing were made using a noise band vocoder in one ear and a low pass filtered signal in the other ear. This is a model for a bimodal fitting with a CI input in one ear and an acoustic (HA)

input in the other severely hearing impaired ear. We assessed localization[2] performance with and without amplification of ILD cues in the low pass acoustic signal using a custom ILD amplification algorithm described in section 5.4.1.

Simulation of bimodal hearing

To simulate the amount of spectral information that would be perceived by a CI user in optimal circumstances, a noise band vocoder (Shannon et al., 1995) was used to reduce the spectral information in the signal. It does not provide a model for loudness perception using a CI. However, with proper settings of the CI and HA signal processing, loudness growth between electric and acoustic stimulation is linear (Francart et al., 2008a), such that for the current purpose of demonstrating improvement of localization performance by ILD amplification, it suffices to have linear loudness growth between the simulated electric and acoustic stimulus.

The noise band vocoder mimics the behavior of a typical CI speech processor by sending the input signal through an (analysis) filter bank, performing envelope detection in each channel and finally adding together noise bands with different frequency contents after modulating them with the corresponding envelopes. Eight channels were used. The analysis filter bank consisted of 4th order Butterworth filters, geometrically distributed between 200 Hz and 7000 Hz. Their frequency responses are shown in the upper panel of figure 5.7. Envelope detection was done by low pass filtering the half wave rectified signal in each band with a cutoff frequency of 300 Hz. The noise bands were generated by filtering a white noise signal with 4th order Butterworth filters whose center frequencies were determined using the Greenwood equation (Greenwood, 1990), which relates position in the cochlea to stimulation frequency. The resulting filters were distributed between 500 Hz and 7000 Hz and are shown in the lower panel of figure 5.7. Note that there is a mismatch between the analysis and synthesis filters. This simulates the mismatch that is for many CI subjects present due to non-individualized filters in the analysis filterbank (see section 1.4.2).

Severe hearing loss was simulated in the contralateral ear by the use of a 6th order low pass Butterworth filter with a cutoff frequency of 500 Hz.

[2]The difference between localization and lateralization is that in the former case the sound image is externalized while in the latter case it is not (Plenge, 1974). In several conditions of our experiments, and especially in experiment 2, it is questionable whether we are still dealing with localization. However, for practical reasons, the term localization is used.

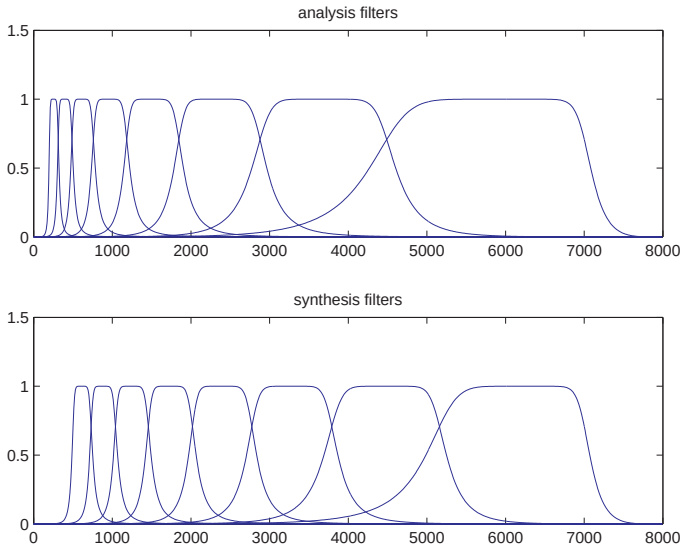


Figure 5.7: Analysis and synthesis filters used for the noise band vocoder CI simulation

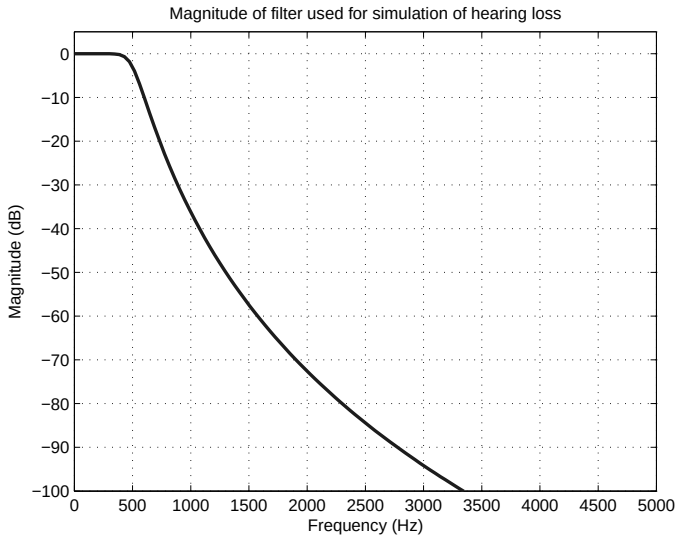


Figure 5.8: Sixth order Butterworth filter used to simulate severe hearing loss

This filter is shown in figure 5.8. The signal was calibrated at 65 dB A, such that for our NH subjects the frequencies up to 500 Hz were clearly audible, 1000 Hz was just audible and higher frequencies were inaudible. This filter simulates an average bimodal system user from our clinic, as for most of these patients the frequencies below $500 - 1000$ Hz can be sufficiently amplified by a HA.

ILD amplification algorithm

A CI stimulates a broad frequency range and if the compression and automatic gain control of the speech processor are optimal, loudness growth can be linear with acoustic loudness growth. Therefore, even though spectral cues are degraded, when considered over a broad frequency range, the head shadow effect is effective for electric stimulation.

For acoustic stimulation in CI users with residual hearing, the situation is different, because their usable residual hearing is mostly limited to about 1000 Hz and the head shadow effect is only physically present for wavelengths shorter than the main dimensions of the head, i.e., frequencies higher than about 1500 Hz (Moore, 2003). Figure 5.6 illustrates this by

showing ILDs for each angle of incidence as determined from the HRTF recordings made for the current study, as described in section 5.2.1.

Therefore an ILD amplification algorithm was developed that makes use of the full-band signals from the microphones of both the HA and CI speech processor. The ILD is determined from these signals and introduced into the low-frequency signal to be emitted by the HA. If A_{CI} is the root mean square (RMS) amplitude of the signal at the microphone of the CI speech processor and A_{HA} the RMS amplitude of the signal at the microphone of the HA, then the ILD in dB is defined as

$$ILD = 20\log(A_{CI}) - 20\log(A_{HA}) \tag{5.2}$$

The ILD is then introduced into the acoustical signal by amplifying it by $ILD/2$.

Note that if the subject has more residual hearing than in the current simulations, it can be useful to amplify only the low frequencies (e.g., using a shelving filter) instead of amplifying the entire frequency range of the acoustic signal.

The effect of our simulations and the ILD amplification algorithm is illustrated in figure 5.9. The "before sim" lines show the levels of the unprocessed signals at the left and right ears for different angles. Around 0° the ILD (difference in level between the two ears) varies approximately linearly with angle while at larger angles the curve flattens. The "vocoder L" and "LP filter R" lines show the levels of the signals at the two ears after a simulation of bimodal hearing. For the left ear (L) a noise band vocoder was used and for the right ear (R) a low pass filter was used (cf. section 5.4.1). The curve for the left (CI) ear remains approximately the same before and after simulation, but the curve for the acoustical signal is severely flattened because of limited ILD cues at low frequencies.

The "LP filter + amp R" line in figure 5.9 shows the same low pass filtered acoustic signal as before, but now the ILD amplification algorithm was applied before the simulation of bimodal hearing. The overall ILD after processing is now as prominent as in the "before sim" stimuli.

Stimuli and conditions

For experiment 2, the two broadband signals were selected from the list: noise14000 and telephone. The noise250 and noise500 signals do not contain large ILDs because they do not have energy at frequencies above 1500 Hz. Therefore their ILDs cannot be amplified by the algorithm. In section 5.5, improvements to the current algorithm are suggested that

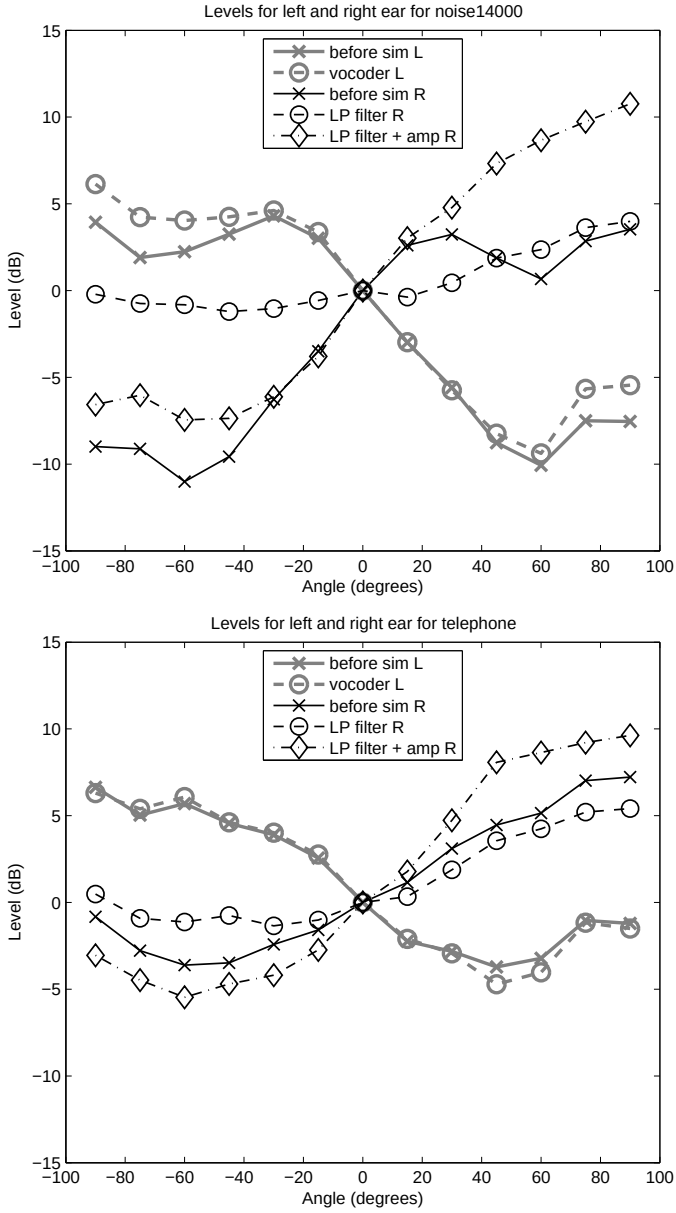


Figure 5.9: Levels of the wide band signals (noise14000 and telephone) after filtering with BTE HRTFs, with and without simulation of bimodal hearing, before and after application of the ILD amplification algorithm. The noise band vocoder simulation (CI) was done for the left ear and the low pass filtering (HA) for the right ear. The ILD at a certain frequency can be obtained by subtracting the respective levels in dB for the left and right ears.

would make it also useful for low frequency signals. The noise3150 signal was not selected since it cannot be perceived using the acoustically stimulated ear.

The stimuli for experiment 2 were created by filtering the signal with HRTFs recorded in a BTE configuration and subsequently simulating bimodal hearing as described in section 5.4.1. There were two conditions, referred to as noamp and amp. In the noamp condition, the simulations were presented without any further processing. In the amp condition, the ILD amplification algorithm as described in section 5.4.1 was applied before the simulation of bimodal hearing.

Procedure

In general, the procedure for experiment 2 was the same as for experiment 1, but as the stimuli were not externalized, the subjects could only respond with angles in the frontal hemisphere (numbers 1-13). The comparison between the results from the two experiments remains valid since in experiment 1 front-back confusions were resolved. Moreover, pilot testing with two subjects did not show any significant differences between a condition in which the subjects could respond at angles in the full horizontal plane or a condition in which they could only respond at angles at the frontal half of the horizontal plane.

As the ILD amplification algorithm introduces artificial ILD cues, subjects needed some training before being able to associate the ILD cues with the correct angles. Therefore, for each combination of stimulus and condition some training runs were performed. A training run was the same as a normal run, but after the subject's response, feedback was shown: it was indicated whether the response was correct or not and the correct response was shown. At least three training runs were done for each stimulus/condition before performing a normal run. Then training runs and normal runs were alternated. Only results for the normal runs were included in the reported results.

Calibration was done separately for each stimulus, using the stimulus from angle 0°. Each channel was calibrated separately such that the level was 65 dB A. This resulted in a stimulus that was approximately balanced in loudness at 0°. This reflects a CI and HA fitting strategy in which the two devices are balanced for a stimulus in front of the subject.

Unlike in real-life situations, in the laboratory setup the sound level of each stimulus was fixed. Therefore monaural loudness cues stemming from the head shadow effect could be used to localize stimuli. As we were interested in the change in localization error associated with the ampli-

fication of binaural cues, these monaural level cues should be reduced to avoid them obscuring effects stemming from binaural cues. Simulations showed that to reduce localization performance using only monaural level cues to chance level, a roving range of R = ±25 dB is required. Such large roving ranges are not feasible due to issues with audibility and uncomfortable loudness. Therefore, as a compromise, during every run, uniform level roving of ±6 dB was introduced. This rove does not completely eliminate monaural loudness cues but it reduces them such that they do not completely obscure differences in localization performance stemming from binaural effects. The effect of changing the roving range is analyzed in appendix C.

5.4.2 Results

The results of experiment 2 are shown in figure 5.10. A repeated measures ANOVA with factors condition and stimulus indicated a significant increase in performance for ILD amplification ($F(1, 81) = 36.7, p < 0.001$).

5.4.3 Discussion

For both stimuli, the conditions without ILD amplification (noamp) yielded worse performance than for any stimulus in the VF-ILDonly condition of experiment 1. This is due to the reduction in spectral detail by the noise band vocoder and the bandwidth restriction by the low pass filter.

For the telephone stimulus, there was a non-significant tendency of improvement in localization performance after amplification of ILD cues by 2° RMS error. For the noise stimulus, performance improved significantly by 14° RMS error. The smaller increase in performance for the telephone signal is probably due the fact that (1) signals with clear modulations merge better between the ears (Francart et al., 2008b), increasing performance in the condition without ILD amplification, (2) the telephone signal had less low frequency content and (3) in the telephone signal, the ILDs available at higher frequencies were smaller than those in the noise1400 signal, and therefore smaller ILDs were introduced in the amp condition, resulting in lower performance.

For both stimuli, the results after ILD amplification were comparable to the results in conditions VF-ILDonly and VF-ILDonly-RP of experiment 1. This means that the ILD amplification algorithm restores localization to the level that is possible with only natural ILDs available.

The bias (not shown) was very small for each condition and stimulus. This is due to the dBA calibration of the stimulus at 0° and the subjects'

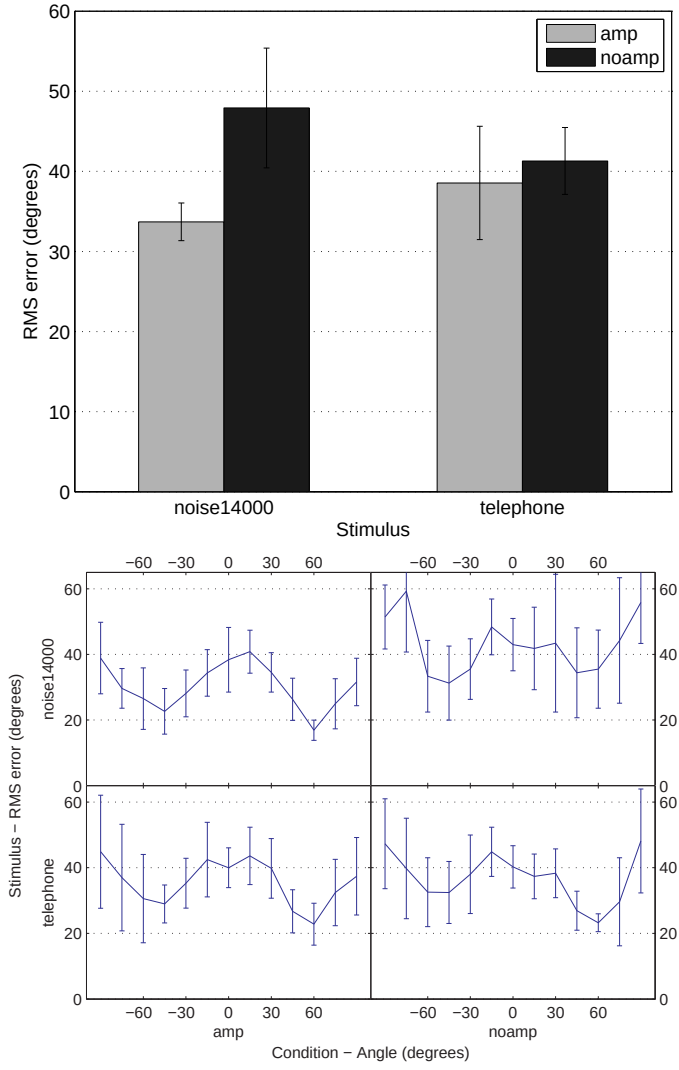


Figure 5.10: The top panel shows the results of experiment 2, averaged for each signal and condition. amp is the condition with application of the ILD amplification algorithm and noamp is the condition without ILD amplification. The bottom panel shows the same results per angle of incidence.

training.

5.5 General discussion and conclusions

While Wightman and Kistler (1992) have shown that ITD is the dominant cue for localization if contradictory cues are available, our data show that if ITDs are not available, ILDs can be as useful for localization as ITDs if the stimulus contains sufficient high frequencies and no conflicting ITD cues are present. This is an important result for users of bilateral CIs and users of bilateral bimodal systems for whom ITDs cues are not available using current clinical signal processing systems.

When simulating bimodal hearing, performance decreased, compared to the condition where only ILDs were available. This was due to the absence of the head shadow effect at low frequencies. When introducing ILDs determined by the high frequencies into the low frequency signal, performance improved by up to 14° RMS error relative to 48° RMS error. This demonstrates the use of a practical ILD amplification algorithm in NH subjects. While the use of a similar algorithm still needs to be tested with CI and HA users and while aspects of combined fitting of the two devices are still to be considered, the current results demonstrate that it is perceptually feasible to use amplified ILDs at low frequencies.

A possible improvement to the current ILD amplification algorithm is to amplify ILDs to larger values than naturally available and introduce them both in the CI and HA signals instead of only introducing them at low frequencies. This may further improve localization performance, which is necessary because ITD cues are still unavailable, both for bilateral CI users and for users of a bilateral bimodal hearing system. Another improvement could be to determine the location of the most prominent sound source using a signal processing algorithm and use an internal mapping function from location to ILD to introduce an unambiguous and sufficiently large ILD into the signal.

Chapter 6

Perception of interaural time differences with bilateral bimodal stimulation

In chapter 4 sensitivity of bimodal listeners to interaural level differences (ILDs) was established. Sensitivity to ILDs could be expected because if the specialized auditory centers could not be used for the detection of ILDs, the task could still be done by comparing loudnesses between the ears at a cognitive level. Sensitivity to interaural time differences (ITDs) is less straightforward because the detection of ITDs is based on binaural correlation and on a microsecond or even millisecond time scale, ITDs cannot be detected at a cognitive level.

Abstract

Sensitivity to ITD was measured in 8 users of a cochlear implant (CI) in the one ear and a hearing aid (HA) in the other severely impaired ear. The stimulus consisted of an electric pulse train of 100 pps and an acoustic filtered click train. Just noticeable differences in ITD were measured using a lateralization paradigm. Four subjects exhibited JNDs in ITD of 156, 341, 254 and 91 µs. The other subjects could not lateralize the stimuli consistently. Only the subjects who could lateralize had average acoustic hearing thresholds at 1000 and 2000 Hz better than 100 dB SPL. The electric signal had to be delayed by on average 1.5 ms to achieve synchronous stimulation at the two auditory nerves.

This chapter is organized in sections introduction (6.1), methods (6.2), results (6.3), discussion (6.4) and conclusions (6.5).

6.1 Introduction

As reviewed in section 1.6.5, changes in ITD of about 10 µs can be detected by normal hearing (NH) subjects in low frequency sinusoids (Yost, 1974). Above 1500 Hz this process breaks down (Yost et al., 1971). While performance with amplitude modulated high frequency sinusoids is worse than with pure tones, performance with so-called transposed stimuli is nearly at the same level as with pure tones (Bernstein and Trahiotis, 2002). Transposed stimuli are generated by modulating a high frequency carrier with a half wave rectified envelope, resulting in similar output of the auditory filters as for the corresponding low-frequency signal.

While it is suggested that listeners are not able to use envelope ITDs in low-frequency sounds, Bernstein and Trahiotis (1985a) showed that those ITDs did affect the lateral position of low-frequency targets. They suggested that the envelopes are seemingly undetectable because the fine structure ITD was dominant at low frequencies.

When the carriers of modulated signals are interaurally discrepant, just noticeable differences in ITD can still be measured, but performance breaks down rapidly with increasing interaural frequency difference (Nuetzel and Hafter, 1981). Blanks et al. (2007) however showed that for a simpler psychophysical task or an animal model, there was sensitivity to ITD when the interaural frequency difference increased up to several octaves.

For many signals both ITD and ILD cues are available and at first sight they appear to be interchangeable. However, Hafter and Carrier (1972) showed that they do yield different percepts and are thus not entirely interchangeable. The just noticeable difference (JND) in ITD is lowest when the ILD is zero (Domnitz, 1973; Shepard and Colburn, 1976).

Recent studies have shown that users of bilateral cochlear implants (CIs) are sensitive to ITDs, although much less so than NH listeners. Best JNDs reported for pulse trains of about 100 pps are around $100 - 200$ µs and for higher pulse rates JNDs are much higher or unmeasurable. JNDs in ITD were measured either through clinical speech processors (Laback et al., 2004; Senn et al., 2005) or with direct computer-controlled stimulation (Laback et al., 2007; Lawson et al., 1998; Long et al., 2003; Majdak et al., 2006; van Hoesel, 2004, 2007; van Hoesel and Tyler, 2003).

In chapter 4 we demonstrated that users of bilateral bimodal hearing systems are sensitive to ILDs. In the current study, we focus on ITD perception by users of bimodal hearing systems. By means of a lateralization task using ITD cues, the best JND in ITD was determined and the delay necessary to synchronize the CI and hearing aid (HA) psychoacoustically was derived.

While localization performance improves when adding a HA to a contralateral CI, users of a clinical bimodal hearing system can most probably not perceive ITD cues. This is due to (1) the signal processing in the CI speech processor, (2) the tonotopic mismatch in stimulation between the two ears and (3) differences in processing delay between the two ears. In this chapter, these technical issues were bypassed and the stimuli were optimized as to achieve maximal lateralization performance using ITD cues.

6.2 Methods

6.2.1 Apparatus

The subjects' clinical devices (CI speech processor and HA) were not used in this study. Our test setup consisted of the APEX 3 program and the hardware described in chapter 2. The shapes and synchrony of the electric and acoustic signals were checked using an oscilloscope.

6.2.2 Stimuli

The acoustic and electric signals were always presented simultaneously and both had a duration of 1 s. The electric signal was a train of biphasic pulses with an inter phase gap of 8 µs and a pulse width of 25 µs. The stimulation mode was monopolar, using both extracochlear reference electrodes in parallel (MP1+2). Electrode numbers will be reported from apex to base, the most apical electrode (A) being electrode 1. This electrode is expected to evoke the lowest perceived pitch and the most basal electrode (B) (electrode 22) to evoke the highest perceived pitch. For each subject, measurements were performed using three target electrodes, one at the first quarter of the array, one in the middle and one at the third quarter of the array. Mostly target electrodes 6 (apical), 11 (middle (M)) and 16 (basal) were used. In preliminary tests, electrode 1 was also included, but as the subjects did not show ITD sensitivity with this electrode, it was not included in the final test protocol.

For the main body of results, an electric pulse train of 100pps, combined with an acoustic filtered click train was used, unless reported otherwise. The acoustic filtered click train was generated using Matlab by adding individual harmonics whose frequencies were multiples of 100 Hz. The harmonics were sines that were added in phase. A discrete set of cutoff frequencies was selected for this study, such that the bandwidth of the

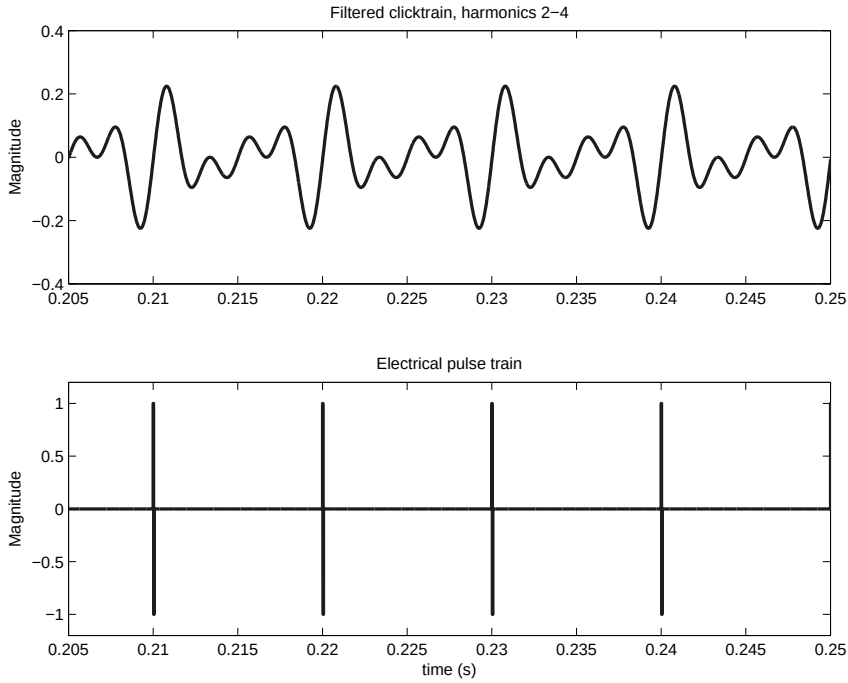


Figure 6.1: Part of an example stimulus. The top panel shows a filtered click train with harmonics 2-4 and $F_0 = 100\,\text{Hz}$ and the bottom panel shows an electric pulse train of 100 pps.

acoustic signal was one octave. The harmonics used for the acoustic signals were 2-4, 4-8, 8-16, 16-32 and 32-64, resulting in cutoff frequencies of $200-400$, $400-800$, $800-1600$, $1600-3200$ and $3200-6400\,\text{Hz}$, respectively. In preliminary experiments, high rate (6300 pps) transposed stimuli (see section 1.6.5) were used as well as click trains of 100 and 150 pps. An example transposed stimulus is shown in figure 6.2.

Example electric and acoustic signals are shown in figure 6.1. In this chapter, the separate electric or acoustic signals are called "signal" and the combination of an electric and acoustic signal is called "stimulus".

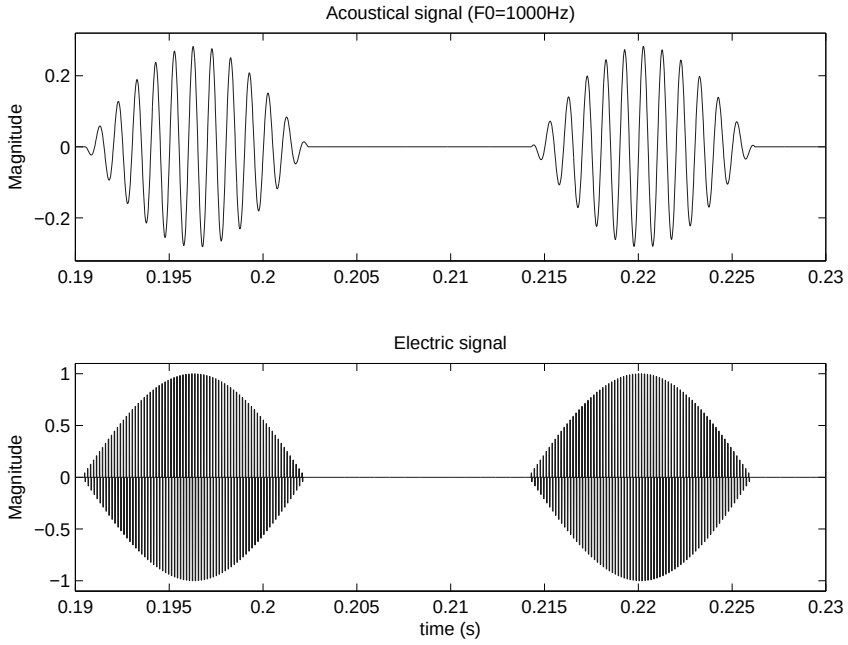


Figure 6.2: Part of an example transposed stimulus. The top panel shows a transposed sinusoid with a base frequency of 1000 Hz and a modulation frequency of 42 Hz. The bottom panel shows an electric pulse train of 6300 pps modulated with a half wave rectified sinusoid with a frequency of 42 Hz.

6.2.3 Procedures

The procedures consisted of four main parts: fitting of T and C levels, stimulus matching, loudness balancing and JND in ITD determination. In the stimulus matching part, the stimulus was determined that yielded the best percept of binaural fusion. Then the loudness between the ears was balanced and finally the JND in ITD was determined using a lateralization paradigm. An overview of the used procedures is shown in figure 6.3.

As ITDs and ILDs both influence the lateralization of a stimulus, great care has to be taken that ITD and ILD are not confused in the procedural design and analysis of the results.

Fitting of T and C levels

During the first test session, the target electrodes (A, M and B) were fitted at 100 and 900 pps. First the hearing threshold (T) was determined and then the comfortable level (C), the loudest level that was not uncomfortable for the subject. At each test session, the fitting of at least one of the electrodes was verified. In all subsequent tests, the electric signal was presented above threshold and the comfortable level was never exceeded.

For every acoustic signal, the hearing threshold and comfortable level were determined at the start of the test session where they were to be used.

Stimulus matching

In NH subjects, envelope ITD perception is optimal if stimulation occurs at approximately the same location in the two cochlea's (Nuetzel and Hafter, 1981). Several methods of matching the place of excitation are reviewed in section 1.5.2. From the ITD perception data, the best match can be derived under the assumption that ITD perception is best for the best-matched stimulus (Nuetzel and Hafter, 1981). As no ITD perception data were available yet, in the stimulus matching phase, one of the target electrodes was combined with each of the target acoustic signals and the subject's task was to indicate which combination yielded the best fused signal. To assist the subjects with their choice, a number of questions were asked about each stimulus and a visual scale of integration was shown and explained (Eddington et al., 2002). This procedure does not yield an exact "match", but gives an indication of it. Finally, a more precise match was determined using ITD perception experiments, assuming that ITD perception is best for the best-matched stimulus.

Depending on the subject's residual hearing, three to five acoustic signals were selected from the list of target signals (see section 6.2.2). The signals with harmonics 2-4, 4-8, and 8-16 were used with all subjects. The other signals were used only if they could be clearly perceived. The signals were subjectively balanced in loudness with the electric signal such that each stimulus was perceived as equally loud and approximately centered. This was done by first selecting a reference electric signal and balancing all acoustic signals with it. Then the subject listened to the different stimuli in random order and had to pick the one that yielded the best binaural fusion. On request the subject could listen to individual stimuli before making a choice.

The resulting stimulus was later on used in the first attempt to lateralize using only ITD cues.

Loudness balancing of stimuli

JNDs in ITD in normal hearing subjects are smallest when the ILD is zero. Balancing of ILD was however complicated by the fact that lateralization is influenced by both ILD and ITD, such that the ITD cannot be set to zero, because the exact delay to be introduced into the electric path is unknown and the ILD cannot be set to zero because the exact balance between the ears is unknown. Therefore, our stimuli were balanced in loudness before assessing the JND.

To avoid differences between monaural and binaural stimulation influencing the results, the signals were presented simultaneously at the two ears during the loudness balancing procedures. Loudness balancing was performed in two steps. In the first step, a loudness balancing experiment was performed with a modified stimulus from which all possible ITD information was removed. In the second step, the balance from step 1 was refined by assessing the extent of lateralization. In subsequent experiments, only the results from step 2 will be used.

The modified stimulus of step 1 was as similar as possible to the target stimulus, but with all possible ITD cues removed. The human auditory system can perceive ITDs in the onset part of a stimulus and in the ongoing part (Laback et al., 2007; Moore, 2003). The onset and offset cue was removed by using a cosinusoidal ramp of 200 ms, yielding a stationary part of 600 ms. The ongoing cue was removed by jittering the time between the individual pulses, both in the electric and the acoustic signal. The degree of jitter introduced was a parameter.

As one subject perceived the jittered electric and acoustic signals as dissimilar, a jitter balancing procedure was first performed to determine

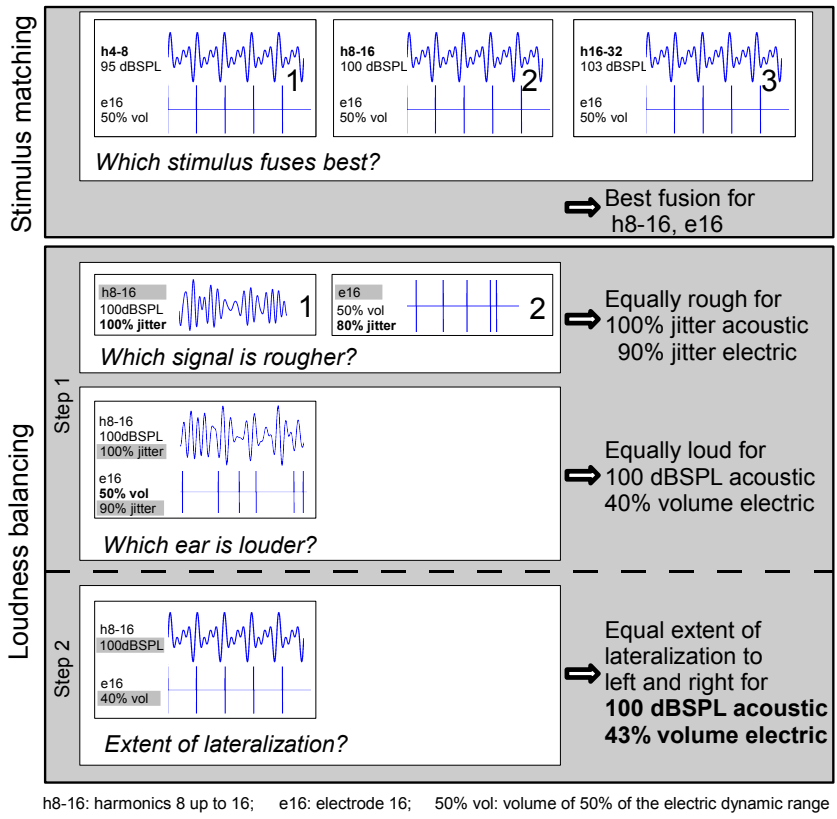


Figure 6.3: Graphical overview of the used fusion and loudness balancing procedures. The white boxes illustrate the stimuli presented to the subject and the text on the right shows example results from each procedure. Each procedure used parameters determined in the previous procedure, which are shown with a gray background. The numbers are fictive but of a realistic magnitude. The plotted electric and acoustic signals only serve illustrative purposes and only show parts of example signals.

the amount of jitter in each signal such that the signals were perceived as similar. The subjects reported that adding jitter to the signals produced a percept of roughness. Therefore a constant stimuli procedure was used in which the acoustic signal was followed by the electric signal and the subject's task was to indicate which signal sounded the most "rough", the first or second. The result was a percentage of acoustic jitter corresponding in roughness to a percentage of electric jitter.

In the subsequent loudness balancing procedure, the acoustic signal was first set to a comfortable level. Then in a constant stimuli procedure the intensity of the electric signal was varied in steps of 10% of the subject's electric dynamic range. The subject was queried for each stimulus whether the sound in the left or right ear sounded louder. A psychometric function was then fitted to the results (Wichmann and Hill, 2001), and the 50% correct point, yielding the electric intensity corresponding in loudness to the acoustic intensity, was determined.

In step 2, the electric intensity from step 1 was refined by assessing whether the sound could be lateralized equally far to the left and right hand side by varying only the ITD. If necessary, a slight change was made to the intensity of the electric signal. The change was always less than 10% of the electric dynamic range. In what follows, only the final result from step 2 will be used. Note that at first, before step 2 could be performed, some familiarization with ITD perception was necessary, ranging from one hour up to multiple sessions.

The result of step 1 and 2 was a stimulus that was balanced in loudness, the equivalent of a stimulus with an ILD of 0 dB in a normal hearing subject. Note that these procedures have to be performed again for each change to either the electric signal or acoustic signal.

Measurement of psychometric functions for ITD

Before determination of the JND in ITD, the subjects were slowly trained by at first presenting them with stimuli with large ITD cues (up to 3 ms off-center) manually and then using a two-alternative forced-choice (2AFC) procedure to assess ITD discrimination. Feedback was never given.

Then, the psychometric function for ITD was determined using a constant stimuli procedure. A number of ITDs was selected over a certain range and a stimulus containing each ITD was presented three times. The subject had to respond whether the sound was lateralized to the left or right side. The ITDs to be presented in one condition were determined manually based on previous subject performance. Some very large ITDs (up to 1.5 ms off-center) were always included to motivate the subject. In

the proximity of the crossover point, the intervals were either 500, 250 or 100 µs, based on the subject's performance.

Psychometric functions were then fitted to the results using the psignifit[1] toolbox version 2.5.6 for Matlab which implements the maximum-likelihood method described by Wichmann and Hill (2001). The 68% confidence intervals around the fitted values were obtained by the BC_A bootstrap method implemented by psignifit, based on 1999 simulations. Results of a psychometric function were only regarded as valid if a confidence interval could be calculated by the bootstrap method and if there was no perfect separation, i.e., if there were points on the slope of the psychometric function different from 0 and 1. If the same experiment was performed multiple times during one test session, the results of those experiments were merged into a single psychometric function. An example psychometric function is shown in figure 6.4.

From each psychometric function, the JND in ITD was determined as half the difference between the 75% point and the 25% point of the psychometric curve. A 68% confidence interval for the JND was determined by combination of the confidence intervals around these points found by the bootstrap method. If multiple psychometric functions were determined for the same condition (e.g., during different test sessions), the median JND was included in the results.

The ITD at the 50% point of the psychometric function indicates the point where the two signals are received synchronously at the auditory nerve. This point corresponds to what would in NH subjects be ITD=0 µs. The acoustic signal travels through the middle ear and part of the inner ear before nerve fibers are stimulated. Therefore, the electric signal has to be delayed. The travel time of the acoustic signal depends on its frequency content: lower frequencies have a larger traveling wave delay in the cochlea. In what follows, this value will be called D_e and is expressed in microseconds delay of the electric signal versus the acoustic signal. Domnitz (1973); Shepard and Colburn (1976) show that in normal hearing subjects JNDs in ITD are smallest when the ILD is zero. Whenever it was unclear what the correct loudness balance was in the current study, the JND in ITD was measured for different balances and the value of D_e for which the JND was the smallest was reported.

[1] see `http://bootstrap-software.org/psignifit/`

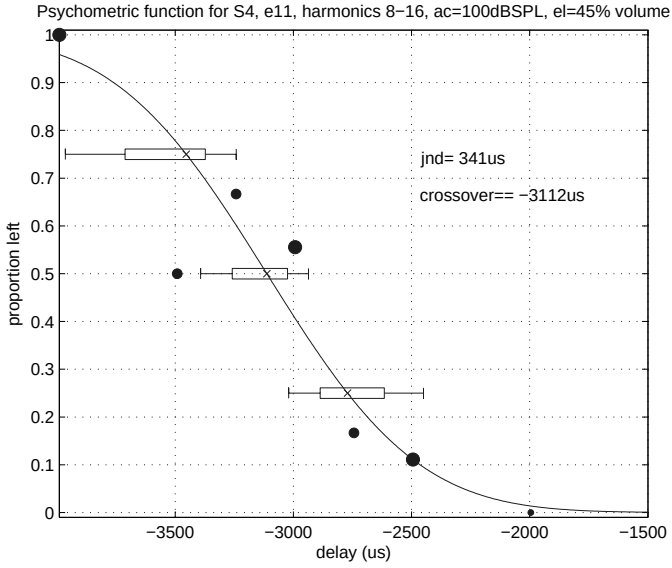


Figure 6.4: Example psychometric function for S4 used to determine the JND in ITD and delay required for psychoacoustically synchronous stimulation. using electrode 11 and harmonics 8-16 for the acoustic signal. The level of the acoustic signal was 100 dB SPL and the level of the electric signal was 45 % of the dynamic range. From the found crossover point (−3112 µs), the delay of the used insert phone (1154 µs) has to be subtracted to find D_e. For the measurement of this psychometric function, 63 trials were used.

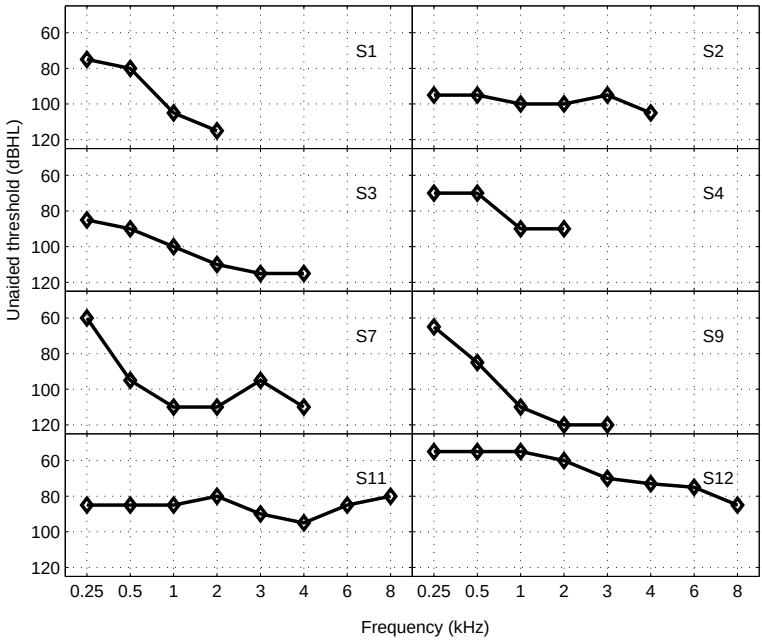


Figure 6.5: Unaided pure tone audiograms for each subject as measured during routine audiometry. Note that the vertical axis starts at 60 dB HL. If no symbol is shown, no threshold could be measured using the clinical audiometry equipment.

6.2.4 Subjects

All subjects were recruited amongst the clinical population of the University Hospital Maastricht (AZM) and the University Hospital Leuven (UZ Gasthuisberg). They were volunteers and signed an informed consent form. This study was approved by the local medical ethical committees. All subjects wore a HA contralaterally to their CI on a daily basis and used a CI of the Nucleus24 type (Cochlear Ltd). S1 and S12 had an electrode array of the Contour Advance type; the other subjects had an array of the Contour type. The clinical processors were of the ESPrit3G type for all but two subjects, who used a Freedom processor instead. All unaided pure tone audiograms as measured during routine audiometry are shown in figure 6.5.

In chapter 4, sensitivity to ILDs was measured in 10 subjects. Of these

Subject	Age (y)	M of use	CI side	Etiology	A	M	B	Perf
S1	58	42	R	Progressive	6	11	16	none
S2	66	49	R	Noise exposure	6	10	16	good
S3	68	58	R	Meniere	6	11	16	none
S4	76	45	R	Progressive	6	11	16	good
S7	40	71	R	Meningitis	6	11	16	poor
S9	32	70	L	Auto-immune	1	6	11	poor
S11	62	25	R	Meniere	6	11	16	good
S12	65	5	L	Genetic (DFNA9)	6	11	16	good

Table 6.1: Subject information: Age is in years at the time of testing. M of use is the number of months of implant use at the time of testing. CI side is left (L) or right (R). The HA was on the other side. Perf is the category of ITD perception performance. A, M and B are the tested electrodes at apical, medial and basal positions in the electrode array.

subjects, 6 were selected for the current study, based on their availability and their ability to perform psychophysical tasks. Two other subjects were included who were implanted more recently (S11 and S12). Relevant data for all 8 participating subjects are shown in table 6.1. The subject numbers used in the current chapter, correspond to those of chapter 4.

The subjects came to the hospital for 2 to 12 test sessions of about 2 hours with at least one week and maximally one month between sessions.

Subject S9 had an incomplete electrode array insertion in the cochlea with two electrodes lying outside of the cochlea. All other subjects had normal electrode insertions.

6.3 Results

6.3.1 Fusion

In preliminary experiments, subjective binaural fusion was assessed for different stimuli (see section 6.2.2). When asked, all subjects except S7 reported that the stimuli with similar envelope fluctuations in the acoustic and electric part yielded binaural integration better than with their own devices (CI and HA). Note that we consider a low rate electric pulse train a signal with a fluctuating envelope.

Subject	JND	Ac	El
S2	$< 25\%$	70%	70%
S4	$< 50\%$	50%	50%
S11	$< 25\%$	100%	100%
S12	?	50%	50%

Table 6.2: JNDs in jitter balancing and percentages of jitter for the acoustic (Ac) and electric (El) signal used for the subsequent balancing experiment.

The subjects were also queried on the similarity in pitch between the two ears. In some cases the stimulus that yielded the best fusion did not yield the best match in pitch. In most subjects, the best fused stimulus yielded a diffuse sound image, i.e., it filled the whole head, but could still be lateralized.

6.3.2 Loudness balancing and intensity

In the first step of loudness balancing, the amount of jitter was determined that yielded the most similar sound between the ears. Adding jitter to a signal added a certain "roughness" to the percept. Therefore in the jitter balancing task, the subjects were asked which signal sounded more rough, the first or the second. The amount of electric jitter that sounded the same as a certain amount of acoustic jitter, was always the same percentage, within the bounds of the subject's JND. Though not measured using a formal procedure, the unilateral JND in jitter was about 10% jitter. Approximate JNDs for bilateral comparisons of the amount of jitter, determined from the psychometric function of the jitter balancing procedure, are given in table 6.2. Although the procedure was not optimal for determining JNDs in jitter discrimination, subjectively they were around 10% jitter for bilateral comparisons. The amount of jitter used for the subsequent loudness balancing task is also listed in table 6.2. The amount of jitter in the electric signal was set to the subject's preference and the corresponding amount of jitter in the acoustic signal was determined based on the jitter balancing procedure. As subject S12 had problems comparing amounts of jitter between the ears, 50% jitter was used in both ears.

The loudness balancing task with the jittered stimulus was at first somewhat confusing for the subjects because the signals could not easily be

lateralized. The subjects had to consciously pay attention to loudness differences between the ears. As a result, performance was somewhat lower than on the loudness balancing tasks in our previous study (Francart et al., 2008a).

In step 2, the results from step 1 were applied to the non-jittered stimulus and the extent of lateralization for large ITD values was assessed. If the image could only be steered to one side (left or right) using only ITDs, the amplitude of the electric signal was adjusted such that the maximal extent of lateralization to each side was symmetric.

When repeating the balancing experiments within the same session, the balancing results were quasi identical. However between sessions differences were observed, mostly in the order of 5 or 10% of the electric dynamic range. Possibly this was correlated with temporary threshold shifts in the residual hearing.

For all subjects except S2, the acoustic signals could be set to a comfortable level. S2 rated the maximal output level of the transducer as "too soft" and no sensitivity to ITD could be observed with the soft signals. Therefore the bandwidth of the acoustic signals was halved, for example using harmonics 16-23 and 23-32 instead of 16-32, yielding a maximal output level that was comfortably loud.

6.3.3 JND in ITD

After balancing the stimuli in loudness, the JND in ITD was assessed using a lateralization procedure. None of the subjects could at first perceive differences in ITD, even when using stimuli that proved successful later on. This is not surprising as these ITD cues are probably not available with their own speech processor in combination with their HA. After training, they reported hearing the stimulus at the back of the head, where it shifted to the left or right according to the ITD.

In several cases, at the beginning of a new test session, subjects could not consistently lateralize stimuli which they could lateralize during the previous test session. Therefore, at the beginning of every new test session, the subjects were trained by presenting them stimuli with large ITD cues. It seems that they did not have a frame of reference for ITD cues and had to be "recalibrated" every test session.

In preliminary experiments, perception of ITDs could be achieved using the stimulus containing transposed signals and the stimulus containing an electric pulse train and an acoustic filtered click train. Of the latter, both 100 pps and 150 pps were assessed. As subject S4 could not perceive ITDs with the 150 pps stimulus, this condition was not included in any

Subject	A	M	B
S2	8-11	11-16	23-32
S4	8-16	8-16	8-16
S7	N/A	N/A	8-16
S9	N/A	8-16	N/A
S11	N/A	8-16/16-32	8-16/16-32
S12	4-8	4-8	32-64

Table 6.3: Range of harmonics of best matching acoustic signal for each electrode and subject. If there was no clear difference between two acoustic signals, both are given.

further tests. As the bandwidth of the filtered click train can be varied, and 100 pps click trains are used in many ITD-CI studies, only the 100 pps click train was used for the final experiments.

Consistent estimates of the JND in ITD could be collected for 4 out of 8 subjects. In figure 6.6 JNDs are reported for each subject for each combination of the acoustic and electric signals that yielded valid JNDs. Each box corresponds to a condition, i.e., a combination of electric and acoustic signals, and the value reported is the median JND in µs that was found for that condition over the different test sessions. If the JND in ITD could not be measured due to insufficient sensitivity to ITD, the condition is marked with a cross.

Figure 6.7 shows the best JND in ITD for each subject and electrode for the acoustic signal that yielded the lowest JND in ITD. Assuming that ITD perception is best for signals matched in place in the cochlea, the figure therefore shows the JND in ITD for place matched stimulation. For comparison, reference values from the bilateral CI literature for pulse trains of 100 pps are given above the label "2x CI". Additionally, table 6.3 lists the acoustic signal for which performance was best for each electrode. In total, 87 psychometric functions were determined for which performance was better than chance level. Each of them was determined based on between 21 and 117 trials.

It should be noted that while the best reported median JNDs for each subject are in the order of $100 - 200$ µs, JNDs were measured as low as 57, 91, 155 and 91 µs for subjects S2, S4, S11 and S12, respectively.

For each subject, we determined whether their ITD perception performance was *none*, *poor* or *good*. Subjects in category *none* could not detect

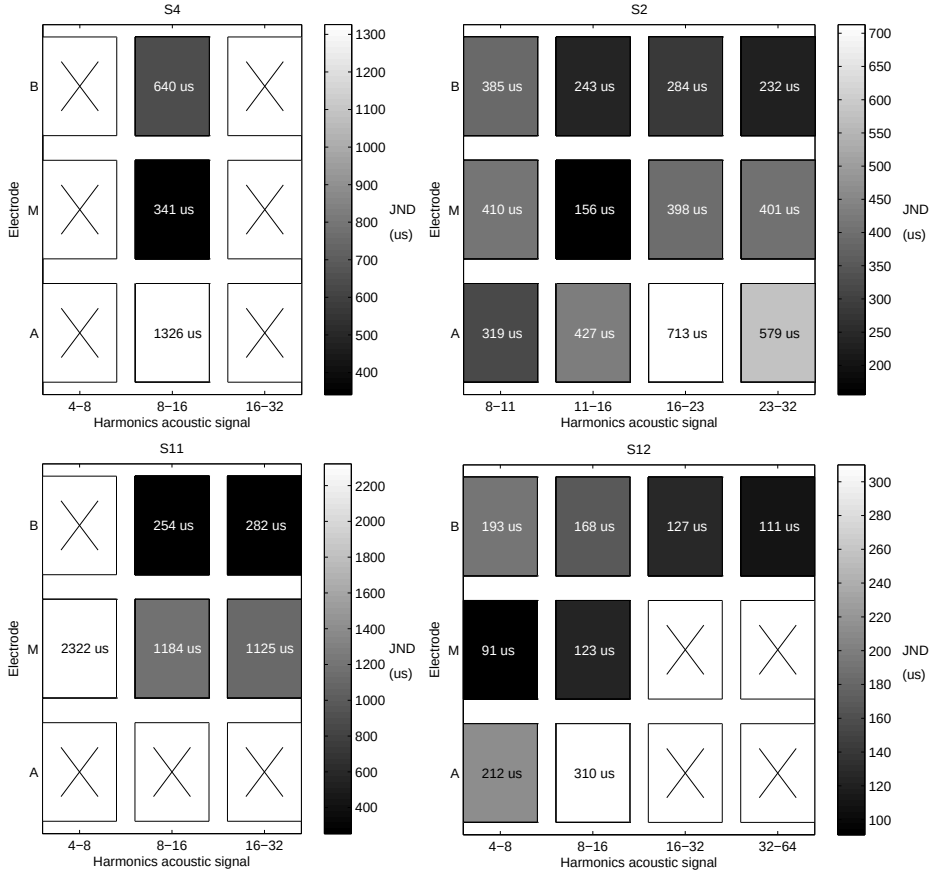


Figure 6.6: JND in ITD in µs for each subject and condition. A cross indicates that the condition was tested, but that sensitivity to ITD was insufficient to do the lateralization task.

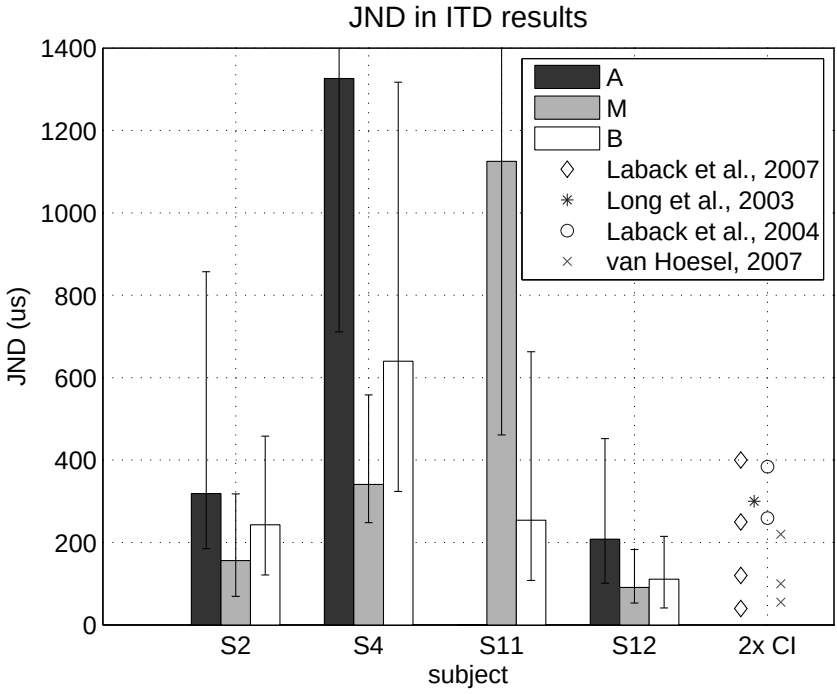


Figure 6.7: Best median JND in ITD per subject and per electrode. The values above the label "2x CI" are reference values from the bilateral CI literature for pulse trains of 100 pps. Each symbol is the JND in ITD for one subject. The error bars are 68% confidence intervals on the fit of the psychometric function, determined by a bootstrap method.

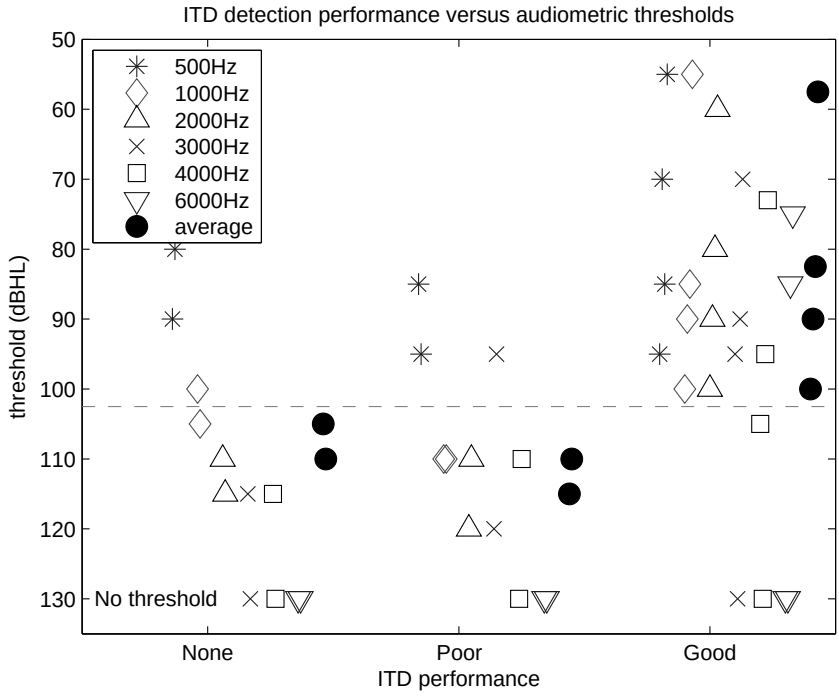


Figure 6.8: ITD perception performance versus thresholds of residual hearing. Each different symbol denotes a different threshold measurement frequency. The filled circles show the average threshold at frequencies 1000 and 2000 Hz.

any ITD at all. Subjects in category *poor* seemed to be able to detect large differences in ITD in informal tests but could not consistently lateralize using only ITD cues. Subjects in category *good* could both detect ITD differences and lateralize using ITDs. In figure 6.8 the three categories are plotted versus the thresholds of the residual hearing of each subject. Whether a subject is in category *good* is related to the average threshold at 1000 and 2000 Hz. A Wilcoxon rank-sum test of difference in threshold between category good and categories none/poor showed a significant effect ($W = 0, p = 0.03$).

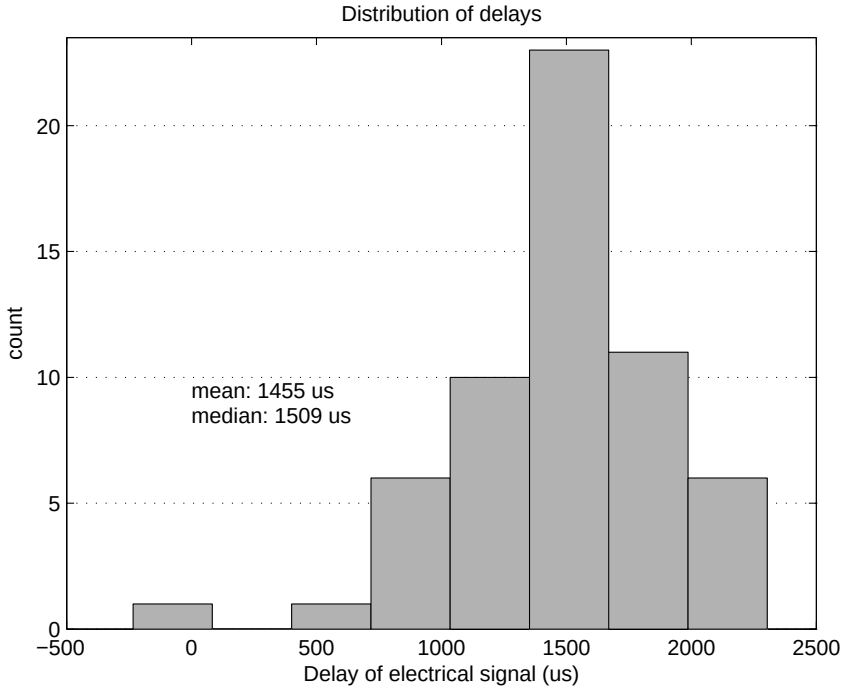


Figure 6.9: Histogram of D_e values. Each value contributing to an increment of one on the vertical axis corresponds to a value found by fitting a psychometric function to the response for between 21 and 117 trials. If measurements with the same stimulus were available at different ILDs, only the D_e was selected for which the corresponding JND in ITD was smallest.

6.3.4 Delays

In addition to the JND in ITD, the psychometric curve also indicates D_e, the point where the two signals are received synchronously at the auditory nerve (see section 6.2.3). Figure 6.9 shows a histogram of delays encountered in all experiments. The median is 1.5 ms. Our data show that D_e depends on the ILD. When ITD perception performance was low, it was entirely disrupted by introduction of a non zero ILD. When performance was high, D_e changed such that the ear with the louder signal had to be delayed relative to the other ear for the stimulus to be centered perceptually. When the amplitude of the acoustic signal was held constant and the amplitude of the electric signal was increased, D_e also increased.

Because the traveling wave delay in the cochlea increases with decreasing frequency, one would expect D_e to vary with the frequency content of the acoustic signal. However, when changing the acoustic signal, the balancing procedure had to be repeated, yielding, possibly a slightly different balance, which influenced D_e and thus confounded the comparison between different stimuli. A clear tendency of change in D_e with changing frequency content was not observed in our data due to (1) balancing differences, (2) the largest possible change in frequency content being severely limited by the amount of residual hearing and (3) the subjects not being sensitive to ITD using the lower electrodes of the CI.

6.3.5 Matching the place of excitation

Assuming that ITD perception is best for signals bilaterally matched in place of excitation in the cochlea, (Nuetzel and Hafter, 1981), the best match in place can be determined by considering the minimum JND in ITD for several acoustic frequencies. For subjects S4 and S11, there were not enough data available, but consideration of figure 6.6 for subjects S2 and S12 reveals a tendency of increasing best acoustic frequency with increasing electrode number. For S2, for electrode 6 performance was best for harmonics 8-11, for electrode 10 for harmonics 11-16 and for electrode 16 for harmonics 23-32. For S12, for electrode 6 performance was best for harmonics 4-8, for electrode 11 for harmonics 4-8 (but with a lower JND than for electrode 6) and for electrode 16 for harmonics 32-64.

6.4 Discussion

6.4.1 Fusion

Although the subjects did respond consistently to the questions in the fusion experiment, this experiment did not always yield the stimulus that was later found to yield the best ITD perception performance. It was, however, useful for preliminary matching of signals and as a training experiment. By varying a parameter of the stimulus and querying for perceptual differences, the subject learns to listen to subtle differences in sound quality and location and learns to describe properties of the percept of a stimulus in a consistent way.

6.4.2 Influence of ILD

The sensitivity to ILD of users of a bimodal hearing system approaches that of normal hearing listeners (Francart et al., 2008a), but their dynamic range is much smaller. Therefore, small differences in ILD can have large perceptual consequences. ITD discrimination in NH subjects is optimal when the ILD is zero (Domnitz, 1973; Shepard and Colburn, 1976). Therefore, in the current study, loudness differences between the ears were eliminated as much as possible. Whenever determination of the JND in ITD was attempted with levels deviating from those according to the results of the loudness balancing procedures, the measurement did not succeed or yielded a large JND. An adjustment of 5% of the electric dynamic range, corresponding to 1 or 2 clinical current units, often made the difference between being able to lateralize using ITD or not.

6.4.3 JND in ITD

The reported JNDs in ITD are poor in comparison to those for NH listeners (Bernstein and Trahiotis, 2002) and of the same order of magnitude as the values found for bilateral CI users (Laback et al., 2004, 2007; Lawson et al., 1998; Long et al., 2003; Majdak et al., 2006; Senn et al., 2005; van Hoesel, 2004, 2007; van Hoesel and Tyler, 2003). While JNDs in ITD of around $100 - 200\,\mu s$ are poor compared to the best fine structure ITD JND of $10\,\mu s$ found in NH subjects (Yost, 1974), they are comparable to envelope ITD JNDs found in NH subjects for the same rate of the modulator (Bernstein and Trahiotis, 2002). Moreover, it should be noted that the residual hearing of our subjects was rather limited relative to that of many subjects who are nowadays receiving a cochlear implant

and possibly better performance would be achieved with better residual hearing.

Our data demonstrate not only ITD perception capability but also lateralization capability. The sound image was clearly steered to the left or right side when introducing ITDs, after carefully balancing the signals in loudness. This means that if clear ITD cues could be transmitted by the CI and HA, they could be used for localization of sound sources and provide advantages such as binaural unmasking, which is very important for speech perception in noise.

For ITDs to be transmitted by a real CI and HA, the two devices must be carefully balanced in loudness and the cutoff frequencies of the band pass filters corresponding to each electrode must be approximately matched to the corresponding acoustic signal in the other ear.

Most probably only onset and envelope ITD cues were used by the bimodal subjects in this study, considering (1) that JNDs for NH listeners with amplitude modulated (AM) signals are comparable to the values found here (Bernstein and Trahiotis, 2002), (2) that CI users are mainly reported to use envelope cues for ITD perception and could therefore be assumed to perceive ITDs using the neural mechanisms that NH listeners use for envelope ITD perception, and (3) the type of signals used in this study; In preliminary experiments no sensitivity to ITD was found at more apical electrodes or lower acoustic frequencies than reported and in most cases our data showed best ITD perception performance for acoustic signals with cutoff frequencies of $800 - 1600\,\text{Hz}$ and $1600 - 3200\,\text{Hz}$. As ITDs in the fine structure of a signal can only be detected up to about 1.3 kHz (Zwislocki and Feldman, 1956), this indicates perception of ITD in the envelope instead of in the fine structure of the signals. This is probably related to our finding that ITD perception performance of bimodal subjects is related to the average thresholds of their residual hearing at 1000 and 2000 Hz. The reason for the subjects' apparent inability to use fine structure ITD cues is currently unclear.

By means of the binaural interaction component in animal models, Noh et al. (2007) have shown that the binaural auditory system can process combinations of electric and acoustic stimulation across ears. This is confirmed by our finding that users of a bimodal system can detect ITDs.

6.4.4 Delays

The median delay required to be introduced into the electric pathway for psychoacoustically synchronous stimulation was 1.5 ms. This is the first report of transmission delay between electric and acoustic stimula-

Publication	Stimulus		
	apex	mid	base
Shallop et al. (1990)	3.82	3.94	4.20
Abbas and Brown (1991)	3.87	3.90	4.12
	Acoustic click		
Nikiforidis et al. (1993)	5.64		
UZLeuven	man: 5.78, woman: 5.57		

Table 6.4: Wave V latencies (in ms) from different studies on ABR and EABR. All were measured at a comfortably loud level. The last row shows reference values used for the clinical ABR setup in our hospital (UZLeuven).

tion. The value of this delay is comparable to the difference between the delays obtained from the acoustic auditory brain stem response (ABR) and electrical auditory brain stem response (EABR) literature. While the paradigms and experimental setups differ between studies, the reported latencies are similar.

In table 6.4, wave V latencies for different studies are summarized. Don and Eggermont (1978) showed that all frequency regions contributed to the ABR but that the response was dominated by contributions from the 2-3 octaves towards the basal end of the cochlea. Therefore, the values in the "base" column of table 6.4 are compared to the acoustic ABR values. On average, the difference is 1.5 ms, which, given the procedural and presentation level differences between studies, corresponds well to the 1.5 ms latency difference found in the current study.

6.4.5 Relation with localization performance and binaural unmasking

In the current study, all parameters were optimized as to achieve optimal ITD sensitivity. It is therefore unlikely that users of current clinical CIs and HAs would be able to benefit directly from ITD cues, given the problems with current clinical devices enumerated in the introduction.

In preliminary tests, our subjects showed similar ITD sensitivity using transposed stimuli with a high pulse rate (6300 pps) and a low modulation frequency (42 Hz). This is comparable to the situation with clinical

devices where intermediate to high pulse rates are used and signals with slow modulations (such as speech) are presented. Therefore, for the four subjects who showed ITD sensitivity, it might be possible to use interaural timing cues in real-world signals if the CI and HA signal processing and fitting is modified to achieve (1) correct binaural loudness growth (Francart et al., 2008a), (2) correct synchronization and (3) correct matching of the places of excitation in the cochleas.

The perception of binaural timing cues can give rise to improved sound localization performance and more importantly binaural unmasking of speech in background noise (Colburn et al., 2006).

6.5 Conclusions

If the average threshold of the residual hearing at 1000 and 2000 Hz is better than about 100 dB HL, lateralization is possible with ITD cues for subjects using a CI in one ear and a HA in the other. The best median JNDs in ITD were 156, 341, 254 and 91 µs in the four of the eight subjects who could discriminate ITDs. This is comparable to the values found in the literature on bilateral CIs. ITDs could in most cases only be detected for acoustic frequencies above about 1 kHz. This indicates that mainly envelope cues were used. For the acoustic and electric signals to be perceived synchronously, the electric signal should be delayed by 1.5 ms.

Chapter 7

Conclusions and further research

7.1 Conclusions

While users of a bilateral bimodal system have binaural inputs, they perform poorly on localization tasks. This can be due to four main technical reasons: place mismatch, incorrect synchronization, non-linear binaural loudness growth and the removal of fine timing cues by the cochlear implant (CI) signal processing (see section 1.5).

However, it depends on the subject's sensitivity to the basic localization cues whether it is worthwhile to solve the technical issues. Therefore we assessed sensitivity the basic localization cues: interaural level difference (ILD) and interaural time difference (ITD).

7.1.1 ILD sensitivity

As it was unknown what the effect of place mismatch is on the perception of ILD cues, in chapter 3 we assessed whether normal hearing (NH) subjects could perceive ILDs in mismatched signals. We found that ILDs could be perceived in uncorrelated signals with a mismatch up to a whole octave but that ILD detection performance decreased with increasing mismatch.

In chapter 4, we measured the just noticeable difference (JND) in ILD of users of a bimodal system. For pitch-matched signals, the average JND was 1.7 dB and for mismatched signals the average JND was 3.0 dB.

From the loudness balancing experiments used to measure the JND in ILD, loudness growth functions between electric and acoustic stimulation could be calculated. Our data do not contradict the observation found in the literature that in bimodal listeners loudness growth is linear between electric and acoustic stimulation on a μA versus dB scale with the slope dependent on both the electric and acoustic dynamic ranges.

7.1.2 Improving localization by ILD amplification

While users of a bimodal system are sensitive to ILD, their residual hearing is in most cases limited to frequencies up to 1000 Hz. Moreover, the head shadow effect is very small at frequencies lower than 1500 Hz. Therefore they will not have access to useful natural ILD cues for real-life localization.

In chapter 5 an algorithm is described and evaluated that determines the ILD of an incoming signal and introduces it into the low frequencies. In simulations with NH subjects, using a wide band noise signal, localization performance improved by more than 14° RMS error relative to 48° RMS error after application of the ILD amplification algorithm.

The ILD amplification algorithm is not only useful for bimodal stimulation but can also improve localization performance for users of bilateral CIs. While bilateral CI users are sensitive to ITD cues to some degree, they cannot perceive them using their clinical systems due to problems in the CI signal processing that are currently unresolved. As we have shown in chapter 5 that, under certain circumstances, localization with only ILD cues can be comparable to localization with ITD cues, the amplification of ILDs that are naturally present could compensate for the lack of ITD cues.

7.1.3 ITD sensitivity

In chapter 6, we measured JNDs in ITD. Four of the eight subjects could lateralize using only ITD cues with JNDs of around $100 - 200\,\mu s$. ITD detection performance was related to the average pure tone threshold at 1000 and 2000 Hz of the residual hearing.

Based on the assumption that ITD perception is optimal if the place of excitation is matched between the ears, we also suggest the use of the JND in ITD for matching the place of excitation. As ITDs cannot be perceived at a cognitive level, and ITD sensitivity is not a measure as subjective as pitch, the method of matching via ITD detection performance might be preferred.

Another important result from the ITD sensitivity study is the delay necessary to psychoacoustically synchronize the electric and acoustic signals. It was found that an average extra delay of 1.5 ms should be introduced into the electric pathway for the two signals to arrive synchronously at the auditory nerve.

7.1.4 Impact on localization performance and binaural unmasking

While we have not measured performance on real-life localization tasks, the results from chapters 4, 5 and 6 are promising for bilateral bimodal CI-users. As the average JND in both ILD and ITD is far below the maximum size of ILD and ITD cues available in real-life signals (see section 1.6), sensitivity to both cues is good enough to perceive real-world localization cues.

As the detection of ITDs is related to binaural unmasking phenomena (Colburn et al., 2006), we are hopeful that with modified CI signal processing and fitting (see section 7.2.2) the subjects could achieve better speech perception in noise.

7.2 Further research

As we have shown the feasibility of perception of ILD and ITD cues by users of bimodal aids, further research into this new field can be done on two parallel topics. On the one hand, the psychophysical investigations can be extended with different signals and conditions and on the other hand there are technical difficulties to be solved for the clinical devices.

In the next sections we first suggest further psychophysical experiments to be conducted and then focus on possible improvements in the signal processing in CI speech processors and hearing aids (HAs) or future integrated processors. Of course, psychophysical experiments and development of signal processing should be done together in an iterative process.

7.2.1 Further psychophysical research

Extension to real-world signals

Now that sensitivity has been shown to ILD and ITD cues with simple stimuli, sensitivity can be measured with more complex stimuli, which are more similar to the stimuli presented using a clinical system. The number of used electrodes can be increased, the pulse rate can be changed and the acoustic bandwidth can be varied.

Varying the different parameters of the electric and acoustic signal and measuring sensitivity to the binaural cues can yield useful information on the changes to be applied to the CI signal processing to optimize sound similarity between ears and perception of binaural cues with realistic signals (e.g., speech).

BMLD

Long et al. (2006) and Van Deun et al. (2008) have shown that both adult and paediatric users of bilateral CIs can exhibit binaural masking level differences (BMLDs). As BMLDs are related to perception of ITDs, BMLD can be assessed in users of bimodal systems using a similar paradigm. If that is successful, it can be extended to more realistic broadband (multi-electrode) stimuli.

Correlation with localization performance

While the JNDs in ILD and ITD found in this thesis are good enough for the perception of ILD and ITD cues in real-life signals, it has not been proven that they are used by bimodal listeners for localization. To establish this, localization performance can be measured with optimally fitted clinical devices and correlated with JNDs in ILD and ITD.

Pitch matching

The novel place matching method using measurement of the JND in ITD (see chapter 6) should be compared with other methods, such as pitch matching (see chapter 4 and Boex et al. (2006); Dorman et al. (2007b)), contralateral masking (James et al., 2001) and analysis of radiographic information of the implanted cochlea (Cohen et al., 1996).

7.2.2 Further technical research

Optimizing binaural loudness growth

For loudness growth to be the same in the two ears, different things should be changed in the CI and HA signal processing. First, the fitting of the CI must be changed such that the compression and mapping yields approximately linear loudness growth on a dB scale. Second, the transfer functions of the automatic gain controls (AGCs) of the CI and HA should be the same or at least the same in the most important part of the dynamic range. Also, the AGCs should be synchronized (e.g., via a wireless link) to avoid attenuation of ILD cues.

Synchronizing devices

Eventual differences in I/O latency of the CI and HA should be compensated and an extra delay of 1.5 ms should be introduced into the electric path to compensate for the acoustic traveling wave delay (see chapter 6).

After chronic stimulation with such synchronized devices, the necessity of electrode-dependent delays can be assessed.

ILD amplification algorithm

Introduction of ILD cues at low frequencies using an ILD amplification algorithm proved successful in chapter 5. However, this algorithm only functions if sufficiently large ILD cues are available at high frequencies. This could be solved by using signal processing techniques to determine the source location using both ILD and ITD cues at the microphone inputs and mapping the location to an ILD to be introduced in the signals at the two ears.

Appendix A

Automatic testing of speech recognition

The APEX 3 program as described in chapter 2 can be used to administer speech recognition tests. In the current appendix, an algorithm is described that can be implemented in APEX 3 to automatically perform speech recognition tests, i.e., without an experimenter present.

Abstract

Speech reception tests are commonly administered by manually scoring the oral response of the subject. This requires a test supervisor to be continuously present. To avoid the latter, a subject can type the response on a computer keyboard, after which it can be scored automatically. However, spelling errors may then be counted as recognition errors, and this will influence the test results. We demonstrate an autocorrection approach based on two scoring algorithms to cope with spelling errors. The first algorithm deals with sentences and is based on word scores. The second algorithm deals with single words and is based on phoneme scores. Both algorithms were evaluated with a corpus of typed answers based on three different Dutch speech materials. The percentage of differences between automatic and manual scoring was determined, in addition to the mean difference in speech recognition threshold. The sentence correction algorithm performed at a higher accuracy than commonly obtained with these speech materials. The word correction algorithm performed better than the human operator. Both algorithms can be used in practice and allow speech reception tests with open set speech materials over the internet.

After the introduction (section A.1), both the sentence correction algorithm and the word correction algorithm are described (section A.2.1 and A.2.2). In section A.3 both algorithms are evaluated. The last three sections contain general results (section A.4) , discussion (section A.5) and

conclusions (section A.6).

A.1 Introduction

Both in clinical practice and in research, speech recognition tests are widely used to assess performance in patients under varying conditions. While speech recognition tests in silence or at a fixed noise level are easy to conduct, they require that the test supervisor is continuously present and scoring is therefore prone to human errors. Speech recognition tests using an adaptive procedure (Levitt, 1971) or even more complex procedures are harder to conduct manually because interaction is needed to change the signal to noise ratio after each trial.

Human errors are due to plain scoring mistakes by the supervisor, but also to unclear pronunciation by the subject. The latter can be an issue in hearing impaired subjects or subjects with a strong dialect.

Both issues can be addressed by using a computer program to automatically conduct the speech test. Subjects enter their response on a computer keyboard and a computer program evaluates the response and selects the next stimulus to be presented. Implementation of such a simple program is straightforward. However, subjects make typing errors, which affect the test results. Therefore, the computer should take into account the possibility of spelling errors and distinguish between such spelling errors and true recognition errors.

Current automatic word correction research can be divided into three broad classes of increasingly difficult problems: (1) isolated word error detection, (2) isolated word error correction and (3) context-dependent word correction (Kukich, 1992). In the first class, errors are only detected, not corrected, mainly by looking up words or N-grams in a dictionary or frequency table. Presently, this class of problems is mainly solved.

The second class – isolated-word error correction – consists of the generation of correction candidates and the ranking of the candidates. A certain input string has to be compared with many entries in a dictionary and amongst the matches, the best match has to be selected. An overview of practical techniques is given by Navarro (2001) and Kukich (1992).

In the third class, context-dependent word correction, not only every individual word is considered but also the words or even sentences surrounding it. Using different approaches such as language models, the noisy channel model, frequency tables and large corpora, the algorithm can then suggest a correction. Reynaert (2005) reviews such algorithms. Research is still going on for this type of problem and while many solutions

exist to subsets of this problem, the general problem remains unsolved.

Spelling correctors from word processing software typically detect word errors using a dictionary and then suggest a number of possible corrections. They solve problem (1) and part of problem (2). It is clear that this approach is not sufficient for automatic correction of speech recognition tests, because in this case, the error must not only be detected but also automatically corrected, without interaction of the user. However, in the case of speech recognition tests, the difficult problem of context-dependent automatic correction can be simplified by using extra information that is readily available: the user does not type a random sentence, but is trying to repeat the sentence that was presented.

In this paper we describe two algorithms for autocorrection of sentences and single words respectively, in the context of speech recognition tests. Both algorithms are evaluated using a custom corpus of manually corrected speech recognition tests and are compared to a simple algorithm that does not take spelling errors into account.

The use of automated speech recognition tests has only been reported a few times in the literature, e.g., in Stickney et al. (2004, 2005), but autocorrection was never used. However, it has many practical applications, both clinically and in a research environment.

Internet speech recognition tests are currently used for screening large populations for hearing loss. Tests exist for both children[1] and adults[2] and are currently available in Dutch (Smits et al., 2006) and are being developed for Dutch, English, French, German, Polish and Swedish in the European Hearcom[3] project. All of these tests make use of closed set speech materials. The use of the reported autocorrection algorithms allows internet tests to be administered with open set speech materials.

This paper consists of two main sections. In section A.2 the two algorithms are described and in section A.3 the development of a test corpus is described and both algorithms are evaluated using that corpus.

A.2 Description of the algorithms

Two algorithms are described in this section, one for correcting words based on a phoneme score and one for correcting sentences based on a

[1] A Dutch hearing screening test for children is available on http://www.kinderhoortest.nl/

[2] Dutch hearing screening tests for adults are available on http://www.hoortest.nl/ and http://www.oorcheck.nl/

[3] More information on the Hearcom internet hearing screening tests can be found on http://www.hearcom.eu/main/Checkingyourhearing/speechtesttext.html

CVC tests

1. Every phoneme that is repeated correctly results in 1 point
2. A phoneme must be exactly correct, even if the difference is small
3. The phonemes must be repeated in the right order
4. Extra phonemes before or after the correctly repeated phonemes have no influence on the score

Sentence tests

1. Every keyword that is repeated correctly results in 1 point
2. A keyword must be exactly correct, e.g., if the plural form is given when the singular form was expected, the word is considered incorrect.
3. Both parts of verbs that can be split must be repeated correctly for the verb to be scored as correct.

Table A.1: Scoring rules for CVC tests and sentence tests

word score. The scoring rules that were used are given in table A.1. In a word test, a word is considered correct if all phonemes are correct. In a sentence test, a sentence is considered correct if all keywords are correct. Keywords are the words that are important to get the meaning of the sentence (thus excluding articles, etc.). Both for manual and automatic scoring, this method requires a list of keywords per sentence. If keywords are defined, both a keyword score and a sentence score can be determined per sentence.

Our algorithm works based on keywords and thus calculates the sentence score based on the keyword score. If no keywords are defined for a certain speech material, it considers all words as keywords and thus considers a sentence correct only if all keywords are correct. The same method is normally used when manually scoring speech recognition tests.

The speech tests that were used to evaluate the algorithms have been normalized using the same scoring rules as implemented in the algorithms.

A.2.1 The sentence algorithm

General

We consider the case where a subject hears a sentence and then has to type this sentence on the computer keyboard. In what follows, the *user input* is the sentence that the test subject types on the computer keyboard, i.e., the sentence to be corrected. The *gold standard* is the sentence that was presented to the subject.

The algorithm processes two input strings: the user input and the gold standard. A sentence consists of words separated by white space. For each word of the gold standard it is manually indicated whether it is a keyword or not, and whether it is part of a split keyword. Split keywords are keywords that consist of two separate words, but only count as one word when it comes to word score. In English, an example would be "The man wrapped up the package", where "wrapped up" would count as one keyword.

Figure A.1 shows the general structure of the algorithm. In what follows, we briefly describe the different blocks:

Input normalization: The punctuation characters , ; . : are replaced by spaces, all remaining non-alphanumeric characters are removed, all diacritics are removed (e.g., `ä` becomes `a`, `è` becomes `e`), all letters are converted to lower case (e.g., `cApiTAL` becomes `capital`) and multiple sequential white space characters are simplified to a single white space character.

Split into words: The *sentence* is split into *words* using the space character as a delimiter. Possible spacing errors are not corrected in this step.

Space correction: Extra spaces in the middle of a word or missing spaces are corrected using the algorithm described in section A.2.1.

Dictionary check: Each word is checked against a dictionary and the results of this check are stored in memory. For our tests, we used the freely available Dutch OpenTaal dictionary (`http://opentaal.org/`).

Number to text: "Words" that consist of only numbers or of a series of numbers followed by a suffix are converted to text using a language specific number to text algorithm. We used a custom algorithm that was manually verified for all numbers from 0 to 10020. Larger

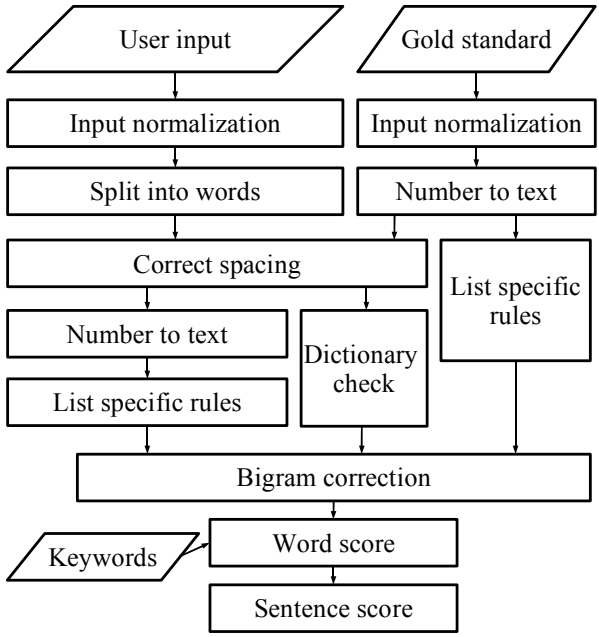


Figure A.1: Flowchart of the sentence algorithm. An arrow signifies that the output from the source block is used as the input for the target block.

1. Replace `cadeau` by `kado`
2. Replace `bureau` by `buro`
3. Replace `eigenaresse` by `eigenares`
4. Replace any number of `d` and `t`'s at the end of a word by a single `t`
5. Replace `ei` by `ij`

Table A.2: Description of regular expressions used for the Dutch LIST and VU sentence test materials

numbers did not occur in the speech materials that were used in the evaluation. While subjects were encouraged to always use the numeric form when typing numbers, this step is still necessary in case they did not follow this rule.

List specific rules: Some language and speech-material specific rules in the form of regular expressions (Friedl, 2006) are applied. If for example the sentence material contains words that can officially be spelled in different ways, one way is selected as the default and the other possibilities are converted to the default. In this stage also some very common spelling mistakes for a language can be corrected. The rules that were used for correction of the Dutch LIST (van Wieringen and Wouters, 2008) and VU (Versfeld et al., 2000) sentence test materials are given in table A.2. These rules are applied to both the user input and the gold standard. Note that the result does not necessarily correspond to the "correct spelling" any more. Therefore, the dictionary check is performed on the data before this transformation.

Bigram correction: The sentence, the results from the dictionary correction and the gold standard are sent to the bigram correction algorithm for the actual spelling correction, as described in section A.2.1.

Word and sentence scores: The word score and sentence score are calculated, as described in section A.2.1.

Spacing correction

The input to the space correction algorithm is a typed sentence that was split into words by using the space character as a delimiter. The algorithm then operates on all unigrams and bigrams that can be formed using these words. A bigram is a combination of any two sequential words in a string. If, for example, the string is "The quick fox jumps", then the bigrams are "The quick", "quick fox" and "fox jumps". Similarly, a single word can be called a unigram. The output of the spacing correction algorithm is a sentence, which is again split into words because the spacing may have changed.

First the basic approximate string matching mechanism is described, then the operation of the entire space correction algorithm is specified. Approximate string matching is done using the concept of anagram hashing (Reynaert, 2004, 2005)[4]. A hash function H is defined that converts a text string W into a numerical value $H(W)$.

$$H(W) = \sum_{i=1}^{|W|} f(w_i)^n \tag{A.1}$$

Here, W is the input string, $|W|$ the length of W and w_1 to $w_{|W|}$ are the characters in W. The function $f(c)$ gives the ASCII value of character c, i.e., a numerical value between 0 and 255. For example, the values for `a`, `z` and the space character are respectively 97, 122 and 32.

Reynaert (2004) found empirically that $n = 5$ is a good value, by considering identical hash function values on very large corpora. To give an idea of the range of $H(w)$, a few examples are given in table A.3.

The hash function is used to compare an input string I with a gold string G. If $H(I) = H(G)$, the strings are assumed to be equal.

It is clear that transpositions of characters can still be present if $H(I) = H(G)$. To allow for character insertions in I, we can iterate over the characters of I and check whether $H(I) - H(i_q) = H(G)$, where i_q is the qth character of I for $1 \leq q \leq |I|$. To allow for character deletions from I, we can iterate over the characters of G and check whether $H(I) + H(g_r) =$

[4]The hashing-part of the space correction algorithm in its current form is similar to calculating an error measure such as the Levenshtein distance (Levenshtein, 1965) between the user input and gold string and allowing for a certain number of errors, but it is faster and easily extensible. A Levenshtein distance calculation algorithm, implemented using dynamic programming has a complexity of $O(|S_1| \cdot |S_2|)$, with $|S_1|$ and $|S_2|$ the length of the input strings S_1 and S_2, whereas the hash-function approach only requires a few hash value calculations, hash table lookups and additions.

Word (w)	$H(w)$
a	8,587,340,257
z	27,027,081,632
aa	17,174,680,514
zz	54,054,163,264
autocorrection	219,976,507,191
the bigram	120,336,090,499

Table A.3: Example values of $H(w)$ for single and double characters, a long word, and a bigram. [From equation A.1 with $N = 5$]

$H(G)$, where g_r is the rth character of G for $1 \leq r \leq |G|$. To allow for character substitutions, we can perform a nested iteration over the characters of I and G and check whether $H(I) - H(i_q) = H(G) - H(g_r)$ with i_q is the qth character of I and g_r the rth character of G for $1 \leq q \leq |I|$ and $1 \leq r \leq |G|$.

Note that $H(A) + H(B) = H(A \oplus B)$, where A and B are strings and $\oplus$ denotes the concatenation operation. We prefer to write $H(A) + H(B)$, because in a real implementation, the values of $H(A)$ and $H(B)$ only have to be calculated once and can then simply be added.

For the space correction algorithm, this check is extended to $H(I) - H(i_q) = H(G) - H(g_r) \pm H(s)$ where s is the space character, i.e., it is checked whether the letters in I correspond to the letters in G with one character replaced by another character and plus or minus the space character.

Now that the above mentioned basic approximate string matching mechanism is clear, we can describe the overall function of the space correction algorithm. It consists of the following steps.

Gold hashing: Every hashed gold word and gold bigram is stored in a hash list.

Bigram checking: The hash value of every user input bigram is checked against the hash list.

Unigram checking: The hash value of every user input word is checked against the hash list.

Space correction: If in the previous steps a matching unigram or bigram is found, it is aligned to the input uni/bigram and spaces from the

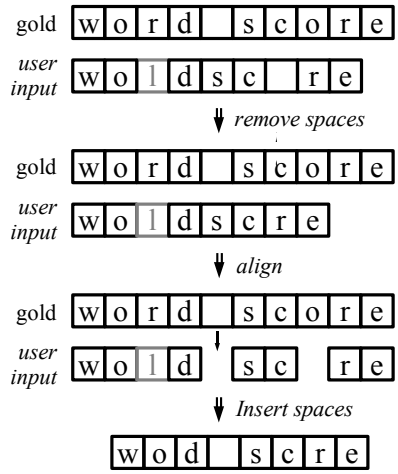


Figure A.2: Example of string alignment. Spaces are marked by empty boxes. In this case the gold string is "word score" and the user input string "woldsc re". First all spaces are removed from the input string. Then both strings are aligned. The space character marked with the single arrow could be inserted into the input string as shown. However, as the percentage of correctly aligned characters ($100 \cdot 7/10 = 70\%$) is smaller than 90%, no space will be inserted because the strings are not considered sufficiently alike in this case.

found uni/bigram are inserted in the input uni/bigram.

Checking against the hash list is done by looking up $H(I) - H(i_q) + H(c_x) \mp H(s)$, with c_x any character from the alphabet, in the hash list and checking if the found value corresponds to I without taking into account any space characters.

When a new bigram is found, it is determined if and where spaces should be inserted. The process is illustrated in figure A.2. First all spaces are removed from the user input. Then it is aligned to the gold standard bigram using a dynamic programming method (Cormen et al., 2001, Chap 15). If the number of corresponding letters is larger than 90%, spaces are inserted in the corresponding places in the user input bigram.

Bigram correction

The bigram correction algorithm takes as input the result from the space correction algorithm that is again split into words using the space character as a delimiter. It operates in the same way as the space correction algorithm, with two differences. The comparison function is now $H(I) - H(i_q) + H(c_x) = H(G)$, i.e., without the extra $\pm H(c_s)$, and words are only considered for correction if they are not in the dictionary.

The result is a corrected list of words in the input sentence that is then sent to the word score determination block.

Word score determination

The word score is calculated by comparing the result from the bigram correction to the gold standard (after the transformations previously described). The score is the number of corrected keywords in the user input that correspond to gold keywords. The corresponding words must occur in the same order in both strings.

To decide if whether two words are the same, the following rules are followed:

- If the user input and gold word are numeric, they must match exactly.
- If the double metaphone (Phillips, 2000) representation of the user input and gold word differ, they are considered different. The double metaphone algorithm (Phillips, 2000) was built for doing phonetic comparisons across languages.
- If the Levenshtein distance (Levenshtein, 1965) between the user input and gold word is larger than 1, they are considered different. The Levenshtein distance is also called the edit distance and is the sum of the minimum number of insertions, deletions and transpositions necessary to transform one string into the other.

Example of the sentence algorithm

We illustrate the function of the entire algorithm by means of an example. Let the user input be

```
Theboy fel from the windaw.
```

and the correct answer

ID	bigram	Hash value
1	the boy	95,540,814,653
2	boy fell	102,798,238,621
3	fell from	113,502,799,457
4	from the	106,245,375,489
5	the window	147,159,170,907

Table A.4: Bigrams and corresponding hash values for "the boy fell from the window"

`The` **boy fell from** `the` **window**

We will use the words in bold as keywords. The user input is transformed by the different functional blocks as follows.

Input normalization: `theboy fel from the windaw`

Correct spacing: `the boy fel from the windaw`

Dictionary check: The words `fel` and `windaw` are not in the dictionary and can thus be corrected.

Bigram correction: The gold bigrams are given in table A.4. The input bigrams that can be corrected (according to the dictionary check) are given in table A.5. Looking up hash values in the hash list and replacing words where appropriate, yields the string: `the boy fell from the window`.

Word score: The gold standard and the corrected input sentence are exactly equal, the word score algorithm yields a keyword score of 4/4 and a corresponding sentence score of 1.

A.2.2 The word algorithm

General

Speech recognition tests can also be done with single words. Typically words with a well defined structure, such as Consonant-Vowel-Consonant (CVC) are used and scores are given based on the number of phonemes identified correctly. In the following sections, an algorithm is described for automated scoring of word-tests.

ID	bigram	Hash value
1	boy fel	88,104,957,853
2	fel from	98,809,518,689
3	the windaw	138,895,929,613

Table A.5: Bigrams and corresponding hash values for bigrams can be corrected from the user input sentence. (a bigram can only be corrected if it contains a word that is not found in the dictionary)

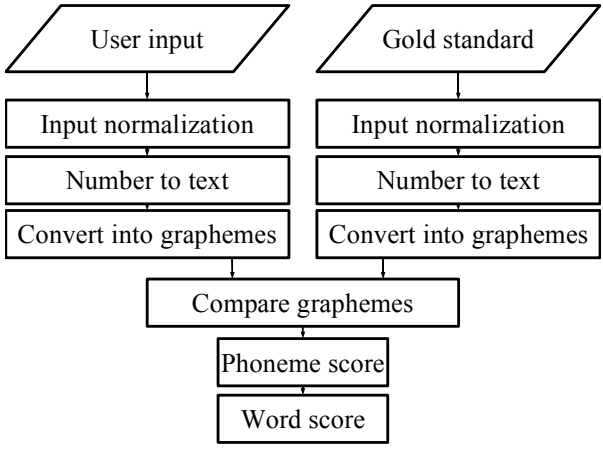


Figure A.3: General structure of the word correction algorithm.

The organization of the word correction algorithm is illustrated in figure A.3, where the main steps are:

Input normalization: The input is transformed to lower case and diacritics[5] and non-alphanumeric characters are removed.

Number to text: If the input consists of only digits, the number is converted to text (using the same number to text algorithm as used in the sentence correction algorithm, section A.2.1).

Conversion into graphemes: The input is converted into a series of grapheme codes (section A.2.2)

Compare graphemes: The user input and gold standard grapheme codes are compared (section A.2.2), resulting in a phoneme score, from which the word score can be derived.

Conversion into graphemes

This module operates on both the user input word and the gold standard word. It makes use of a language-specific list of graphemes. A grapheme is the set of units of a writing system (as letters and letter combinations) that represent a phoneme. Every grapheme corresponds to a numeric grapheme code. Graphemes that sound the same receive the same code. The list that is currently used for Dutch is given in table A.6. Some graphemes correspond to the same phoneme only if they occur at the end of a word and not if they occur in the middle of a word. Therefore, if a `g` or `d` occurs at the end of the word, it is converted to the code of `ch` or `t`.

The algorithm looks for the longest possible grapheme in the string. For example `boot` would be converted into [2 40 19] and not into [2 14 14 19].

Compare graphemes

The phoneme score is calculated by comparing the two arrays of grapheme codes. First, graphemes that do not occur in the user input grapheme list are removed from the gold grapheme list and graphemes that do not occur in the gold grapheme list are removed from the user input grapheme list. Then the score is calculated as the number of corresponding graphemes for the best alignment of both arrays. The best alignment is defined as the alignment that yields the highest score.

[5] Note that if speech materials are used where diacritics influence the correctness of the result, they should not be removed.

[a]$_1$ [b]$_2$ [c]$_3$ [d]$_4$ [e]$_5$ [f]$_6$ [h]$_7$ [i]$_8$ [j]$_9$ [k]$_{10}$ [l]$_{11}$ [m]$_{12}$ [n]$_{13}$ [o]$_{14}$ [p]$_{15}$ [q]$_{16}$ [r]$_{17}$ [s]$_{18}$ [t]$_{19}$ [u]$_{20}$ [v]$_{21}$ [x]$_{22}$ [y]$_{23}$ [z]$_{24}$
[ch]$_{25}$ [g]$_{26}$ [oe]$_{27}$ [ui]$_{28}$ [aa]$_{29}$ [ee]$_{30}$ [ie]$_{31}$ [uu]$_{32}$ [ng]$_{33}$ [ij ei]$_{35}$ [uw w]$_{37}$ [ou au]$_{39}$ [oa oo]$_{40}$

Table A.6: Graphemes used for correction of Dutch CVC words. Graphemes with the same code are between square brackets and codes are given as subscripts.

Example of the word algorithm

As an example, the word `kieuw` is presented to the test subject. If the typed word (user input) is `kiew`, the autocorrection proceeds as follows:

Grapheme conversion: `kieuw` is converted to [10 31 37] and `kiew` is converted to [10 31 37]

Grapheme comparison: Correlation of the 5 different alignment positions of both arrays yields [0 0 3 0 0], thus the score becomes 3

As a second example, the word `dijk` is presented to the test subject. If the typed word (user input) is `bij`, the autocorrection proceeds as follows:

Grapheme conversion: `dijk` is converted to [4 35 10] and `bij` is converted to [2 35]

Grapheme comparison: As the graphemes number 4, 10 and 2 only occur in one of both arrays, they are removed. The resulting arrays are [35] and [35]. Cross correlation of both arrays yields [1], thus the score becomes 1.

A.3 Evaluation of the algorithms

To assess the feasibility of the usage of both algorithms in practice, a corpus of typed responses to speech recognition tests was developed and used to evaluate the algorithms. The difference in score between the autocorrection algorithm and the manual score is determined and compared to the error introduced by mistakes of the operator when manually scoring the speech tests.

A.3.1 Development of a test corpus: procedures

To develop a test corpus, a clinical test setup was reproduced. However, in addition to repeating the speech token they heard, the subjects also typed the token on the keyboard of a computer running the APEX program (Francart et al., 2008e). The operator then scored the oral response using the standard procedures for each speech material. Two final year university students of audiology conducted the experiments, and the analyses described in this paper were performed by a third person.

For each speech material, clear rules were established for obtaining the score, corresponding to the rules that were used for the normalization of the speech materials and described in the corresponding papers. All subject responses and manual corrections as described in the next paragraph were combined into a test corpus that can be used to fine-tune or evaluate an autocorrection algorithm.

A corpus entry consists of the following elements:

Correct sentence: The sentence as it was presented to the subject, annotated with keywords and split keywords

Subject response: The string as it was typed on the computer keyboard by the test subject.

Manual score: (MO) The score that was given by the audiologist using the *oral* response. This was done by indicating correctly repeated words of a subject on a printed copy of the speech token lists and manually calculating the word score and sentence score on the spot.

Corrected manual score: (MOC) In a first iteration, typed responses were run though the autocorrection algorithm and every difference in score between the algorithm and the manual score of the oral response was analyzed by the operator using the notes made during the experiment. If the operator appeared to have made an error while scoring the oral response, it was corrected.

Manual score based on typed response (MT) Every string entered by the subject was manually scored, thereby ignoring spelling errors.

If only the pure autocorrection aspect of the algorithm is evaluated, the MT scores are the most relevant ones. It corresponds to presenting the typed input sentences to a human operator and to the autocorrection algorithm and have both calculate the word score and sentence score.

To assess performance in real-life situations, the MOC scores have to be considered. Here "the perfect operator" is used as a reference. Differences between the MOC and MT scores are due to differences between the oral response and the typed response.

Finally, differences between the MO and MOC scores correspond to errors made by the operator that will be present in any real experiment. Thus, to assess real-life performance of the algorithm, it is useful to consider the difference between the errors made by the operator and errors made by the algorithm, i.e., the extra errors introduced by the algorithm.

A.3.2 Materials

Three different Dutch speech materials were used to develop three different corpora:

NVA words: (Wouters et al., 1994) 15 lists of 12 consonant-vowel-consonant (CVC) words, uttered by a male speaker.

LIST sentences: (van Wieringen and Wouters, 2008) 35 lists of 10 sentences, uttered by a female speaker. Each list contains 32 or 33 keywords. A sentence is considered correct if all keywords are repeated correctly and in the right order. Both a keyword score and sentence score are defined.

VU sentences: (Versfeld et al., 2000) 39 lists of 13 sentences, uttered by a male speaker. A sentence is considered correct if all words, not only keywords, are repeated correctly. Usually, only a sentence score is used with the VU sentences.

The NVA words were presented in quiet at 3 different sound pressure levels (well audible, around 50% performance and below 50% performance, ranging from 20dBSPL up to 65dBSPL).

The LIST and VU sentence materials were masked by four different noise materials: speech shaped noise, a competing speaker in Dutch, a competing speaker in Swedish and the ICRA5-250 speech shaped noise modulated with a speech envelope (Wagener et al., 2006). The operator was instructed to measure at least 3 signal to noise ratios (SNR) for each condition with the purpose of determining the speech reception threshold (SRT) by fitting a psychometric curve through these points afterwards. The number of measured SNRs per condition varied between 3 and 5 and the used SNR varied between -20dB and 10dB. The SNRs used for each subject were recorded.

As the sentence algorithm is based on keywords, we marked keywords for the VU sentences and used these for the algorithm. They were marked according to the same rules that were used for the LIST sentence material (van Wieringen and Wouters, 2008). In simplified form, this means that all words are keywords except pronouns, adpositions, auxiliary verbs and articles. This condition is labeled (word).

In clinical practice however, the VU sentences are scored differently: a sentence is only counted as correct if all words, not only keywords, are repeated correctly. Therefore, we also performed autocorrection with all words in the gold standard sentence as keywords instead of only the keywords that were marked. This condition is labeled (sent).

A.3.3 Subjects

To obtain a diverse corpus, 20 young students of the university of Leuven were recruited, (group 1), as well as 13 subjects who reported having problems with spelling and computer use (group 2), aged 50 years old on average.

A.3.4 Evaluation

We evaluated both autocorrection (Ac) algorithms using our different corpora. First, we measured the number of false positives and negatives. Second, we assessed the influence on the obtained SRT, the value that is traditionally derived from speech recognition tests.

For comparison, we also performed autocorrection on our corpora using a simple algorithm that counts the number of keywords that are exactly the same in the input sentence and in the gold standard, and that occur in the same order. This algorithm is labeled SIMPLE. While it has a very high false negative rate, the results give an impression of the number of spelling mistakes that were made in each condition.

To evaluate of the number of errors made by the human operators, the percentages of modifications between the MO and MOC conditions were calculated.

Percent correct

We calculated the difference between the manual scores (for word score and sentence score) and the automatically generated scores. The difference is given as percentage errors of the autocorrector, with the manual score

Test material	Grp	#	Simple	MO-Ac		MOC-Ac		MT-Ac		MO-MOC		ΔSRT
				Word	Sent	Word	Sent	Word	Sent	Word	Sent	
LIST	1	3280	12.7	1.5	1.3	0.5	0.6	0.4	0.4	1.0	0.7	0.06
LIST	2	1310	26.6	2.3	3.7	1.1	2.2	0.2	0.3	1.3	1.5	0.23
VU (keyw)	1	3354	6.0	1.5	2.1	0.5	0.7	0.3	0.2	1.0	0.6	0.55
VU (keyw)	2	1742	16.5	1.7	3.0	1.0	1.9	0.4	0.4	0.9	1.0	0.18
VU (allw)	1	3354	6.0		3.9		3.3		2.9	1.0	0.6	0.66
VU (allw)	2	1742	16.5		6.3		5.4		3.9	0.9	1.0	0.61
				Phon	Word	Phon	Word	Phon	Word	Phon	Word	
NVA	1	756	32.8	1.3	4.0	0.3	0.9	0.1	0.3	1.0	3.0	
NVA	2	600	38.0	3.4	8.8	1.2	3.5	0.2	0.5	2.3	5.7	

Table A.7: Percentage of errors made by the autocorrection algorithm compared to manual scoring methods for each speech material, group of subjects (Grp), number of tokens in the corpus and corpus entry type. For the sentence materials, errors for keyword score (Word) and for sentence score (Sent) are given. For the CVC materials, errors for phoneme score and for word score are given. # is the total number of sentences presented for the sentence tests and the total number of words presented for the CVC test. MO-MOC is the percentage of changes between the MO and MOC scores. ΔSRT is the mean of the differences in estimated SRT (in dB) between Ac and MOC for each condition. MO is the original manual score based on the oral response, MOC is the corrected manual score based on the oral response, MT is the manual score based on the typed response and Ac is the score by the autocorrection algorithm.

as a reference. As there are different manual scores (cf. section A.3.1), several scores are given for each condition in table A.7.

For sentences, results are given for word score and sentence score. The figures for word score (Word) reflect the number of keywords that were incorrectly scored by the autocorrector per total number of keywords. The sentence score (Sent) is based on the keyword score that is commonly used for the LIST sentences and is in clinical practiçe the only score used for the VU sentences.

For words, results are given for phoneme score and word score. The word score is based on the phoneme score. Here, the phoneme score is the most realistic indicator, as in practice phoneme scores are commonly used.

Influence on SRT

The SRT is commonly determined by fitting a two-parameter logistic function through the percent correct values found at different SNRs recorded during the tests. We assessed the influence of the autocorrection algorithm on the estimated SRT by calculating the difference between the SRT determined by fitting the percent correct values by manual scoring (MOC) and by the autocorrection algorithm (Ac).

There were always three or more data points (SNR values) per condition (speech material/noise type/subject). The average difference in SRT between manual and automatic scoring for each speech material is given in the last column of table A.7. As the accuracy of the SRT determined by this method is usually not better than ±1dB (van Wieringen and Wouters, 2008; Versfeld et al., 2000), our algorithm will have no significant impact on a single estimated SRT value if the difference remains below this value.

A.4 Results

Table A.7 shows percentages of errors of the autocorrection algorithm versus the different manually scored entries in the corpus. In the first column the different speech materials are given. For the VU speech material, the text "(keywords)" or "(all words)" indicates whether all words of the sentence were used as keywords or only the words marked as keywords. For each speech material results are given for both groups of subjects, group 1 are the "good" spellers and group 2 the "bad" spellers. In the second column, the number of tokens in our corpus is given and in the third column the percentage of errors made by the simple algorithm. The next

8 columns give percentages of errors per corpus entry type as described section A.3.1. For each corpus entry type, two parts are given: the results with word scoring (Word) and the results with sentence scoring (Sent). Similar, for the NVA words the results are given for phoneme scoring (Phon) and for word scoring (Word). The last column of the table gives the mean difference in speech reception threshold (SRT) calculated on the Ac and MOC results.

In what follows, we will first give some observations on the number of tokens per test material and group, then we will compare the results between both groups of subjects. Thereafter, we will compare the columns labeled MO-Ac, MOC-Ac, MT-Ac and MO-MOC with each other and finally we will analyze the differences between rows, i.e., between the different test materials and between the (keywords) and (all words) conditions for the VU sentences.

First, considering the number of tokens presented, 3280 (group 1) or 1310 (group 2) LIST sentences correspond to 328 or 131 lists of 10 sentences. This means that overall each of the 35 lists of sentences was presented at least 3 times to each group of subjects and often more. For the VU sentences, similarly, 258 or 134 lists of 13 sentences were presented corresponding to at least 3 presentations to each group of subjects. Similarly, for the NVA words 63 and 50 lists of 12 words were presented. This means that each of the 15 NVA word lists was presented at least 3 times to each group of subjects.

Comparison of the results of group 1 and 2, the "good" and the "bad" spellers, shows that the simple algorithm (column 4) made many more errors with the data of group 2. The results from the simple algorithm are, of course, the same for the VU (keywords) and VU (all words) conditions, but are shown twice for clarity. Comparison of autocorrection performance between the two groups (columns MO-Ac, MOC-Ac and MT-Ac), shows that slightly more errors were made with the data of group 2, on average 0.5 % difference in word score errors for the LIST sentences and 0.3 % for the VU sentences.

In the following paragraphs, we will first compare the percentages of errors between sentence scores (Sent) and words scores (Word), and then compare the results for the different corpus entry types. We will compare the MOC-Ac and MT-Ac scores, followed by the MO-Ac and MOC-Ac scores and then consider the MO-MOC scores. All comparisons will be done per column, i.e., for all test materials and both groups simultaneously.

For the LIST and VU sentence tests, the percentages of errors for the sentence scores (Sent) tend to be somewhat larger than those for the word

scores (Word). This is due to the fact that any word of a sentence that was scored incorrectly leads to an error in the score of the entire sentence, while in the case of word scoring, it only leads to an error for one of the words of the sentence. For the NVA words, the same is true for phoneme scores versus word scores.

The difference between the MOC-Ac scores and MT-Ac scores (columns 7-8 and 9-10) is related to the difference between the typed response and the oral response. It gives an indication of how difficult it was for the subjects to combine the typing task with the speech perception task. The average difference between the MOC-Ac scores and the MT-Ac scores is 0.5 %.

The differences between the MO and MOC scores correspond to errors introduced by manually scoring the speech tests, either by misunderstanding the oral response or by miscalculating the resulting score. Comparison of column 5-6 (MO-Ac) and 7-8 (MO**C**-Ac) shows that both the word scores and sentence scores improve, on average, by 1.0 % less errors of the autocorrection algorithm.

The MO-MOC column indicates purely the number of errors the made by the human operator. The average human error for word scoring of sentences (LIST and VU) is 1.0 % and for sentence scoring it is 0.9 %. Comparison of these values to the values in column MOC-Ac and column MT-Ac, shows that the average number of errors made by the autocorrection algorithm is smaller. For word scoring (Word), the differences between MO-MOC and MOC-Ac/MT-Ac were significant ($p < 0.01$, paired t-tests) for the LIST sentences in both groups, for the VU sentences in group 1 and for phoneme score of the NVA words in both groups.

Now we will consider differences between the rows of the table. Comparison of the autocorrection performance between the LIST and VU sentences reveals no significant difference using a paired t-test for any of both groups of subjects. However, comparison of the scores for the VU sentences with sentence and keyword scoring respectively, shows that the algorithm performs significantly ($p < 0.01$) better with keyword scoring than with sentence scoring for the VU sentences. The reason is that a human operator tends to ignore or just mishears small errors in words that are irrelevant for the meaning of the sentence. For sentence scoring with this speech material, every word is considered a keyword and thus influences the sentence score.

The ΔSRT values in the last column give the mean difference in SRT found from the psychometric function when using the MOC and Ac scores. While all ΔSRT values differ significantly from each other, both between groups and between speech materials, the absolute differences are very

small and there is no clear tendency of change. Note that for each condition only 3 or 4 SNRs were measured and that ΔSRT will decrease if more SNRs are measured per condition. For example for group 1 the percentage of errors for the LIST sentences is 0.6% (MOC-Ac). In the case of 3 SNRs measured per condition, the average number of errors per condition is $10 \times 3 \times 0.006 = 0.18$. This means that in most cases the SRT will not be influenced at all, but if there is an error present in any of the sentences of this condition, it may have a large influence on the SRT because the psychometric curve (with 2 parameters) is fit through only 3 data points.

A.5 Discussion

The within-subjects standard deviation on the SRT determined using an adaptive procedure is 1.17 dB for the LIST sentences in noise (van Wieringen and Wouters, 2008) and 1.07 dB for the VU sentences in noise (Versfeld et al., 2000). The error introduced by using the autocorrection algorithm is an order of magnitude smaller, and will therefore not influence the results of a single SRT measurement.

In order to assess real-life performance of the autocorrection algorithms, the MOC scores should be used as comparison, because these compare the algorithms to a well established standard, i.e., manual scoring of oral responses. When only percent correct scores are considered (no SRT calculation), very small errors are obtained, in most cases even below the expected accuracy of testing. Moreover, the number of errors made by the autocorrection algorithm is similar to or smaller than the number of errors made by the operator, so the results will not be influenced more by the use of autocorrection than by the errors made by a human operator, or by the possible bias of the human operator.

The simple non-autocorrecting algorithm (Simple) should not be used in practice, especially when the subjects are expected to have problems with spelling. Comparison of the results of the simple algorithm between group 1 and group 2 reveals that the subjects of group 2 indeed made many more spelling errors. Comparison of the results of the autocorrection algorithms between groups shows that the results are slightly worse for group 2, but still within the expected accuracy of speech recognition tests. It should, however, be noted that the subjects in group 2 were selected based on their self-reported problems with spelling and computer use. Therefore, the results of group 2 should be observed as worst case results. In normal circumstances, these subjects would probably not be tested using an automated setup because they need a lot of encouragement

and repeated instructions. Nevertheless, the percentages of errors of the autocorrection algorithm are still in the same range as the percentages of errors made by a human operator. The algorithm itself copes very well with this difficult task.

Comparison of the MOC and MT scores shows that there is a small difference between the operator's assessment of the oral response and the typed response. It is, however, never clear what the intended answer is: did the subject intend to answer what he said or what he typed?

The word correction algorithm performs better than the human operator. This is probably due to, on the one hand, unclear articulation of the subjects, and, on the other hand, the difficulty of the task: the experimenter has to remain very well concentrated during the repetitive task and in a few seconds time has to decide which phonemes were repeated correctly. When the answer is approximately correct, there is a chance of positive bias and when the answer is incorrect, it is not always straightforward to identify a single correctly identified phoneme using the strict rules.

Moreover, data of the VU sentences indicate that, while the percentages of errors for sentence scoring are acceptable, the algorithm is best used with keyword scoring. As a human experimenter tends to ignore small errors in words that do not contribute to the meaning of the sentence – even if the response is incorrect according to the strict rules – keyword scoring using the autocorrection algorithm is most similar to this situation.

Both algorithms were developed for the Dutch language. Applicability to other languages depends on the correspondence between phonemes and graphemes of the language. While in Dutch this correspondence is rather strict, this is not necessarily the case in other languages (e.g., English). In any case, we expect the autocorrection algorithms to perform very well in amongst others Danish, Finnish, French, German, Italian, Polish, Spanish, Swedish and Turkish because in these languages the correspondence between phonemes and graphemes is strong. In order to convert the sentence algorithm to another language, the only blocks that have to be changed are the language and speech material specific rules and of course the list of keywords of the speech material. In order to convert the word algorithm to another language, only the list of phonemes and phoneme codes has to be changed.

A.6 Conclusion and applications

The autocorrection algorithms for both sentence tests and word tests are very well suited for use in practice and will not introduce more errors than a human operator.

In a clinical setting, the use of automated speech recognition tests may be rather limited because the test takes longer, the subjects require clear instructions anyway and some subjects may not be able to efficiently use a computer keyboard. However, automated speech recognition tests can be very useful in many other area's, including research, screening large groups of patients and remote tests (e.g., over the internet).

When a test subject does not articulate clearly it can be very difficult to score single words manually, especially when testing hearing impaired subjects. In this case automatic scoring using our autocorrection algorithm should be preferred over manual scoring.

Appendix B

Roving for across frequency ILD perception

In this appendix, the amount of roving necessary for the ILD perception experiment of chapter 3 is calculated.

A standard is presented, followed by a stimulus (see section 3.2.1 on p71). Let S_L, S_R be the level of the standard, left and right; L_L, L_R the level of the stimulus, left and right; I the ILD presented; and R the maximal rove level.

The ILD is introduced as follows:

$$L_L = S_L + \frac{I}{2} \tag{B.1}$$

$$L_R = S_R - \frac{I}{2} \tag{B.2}$$

Let r be the rove for a trial, $r \in [-R, R]$

$$L_L = S_L + \frac{I}{2} + r \tag{B.3}$$

$$L_R = S_R - \frac{I}{2} + r \tag{B.4}$$

We calculate the chance that the subject answers correctly by monitoring only the left ear, i.e., $L_L > S_L$

$$p(L_L > S_L) = \frac{\frac{I}{2} + R}{2R} \tag{B.5}$$

$$= \frac{I}{4R} + \frac{1}{2} \tag{B.6}$$

For a 1up/2down procedure, the chance level $p(L_L > S_L) = P = 0.71$

$$I = 4R(P - 0.5) \tag{B.7}$$

Thus for $R = 5$ we find $I = 4.2$ and for $R = 10$ we find $I = 8.4$. Therefore if we use a rove of R = 10 dB all JNDs must be < 8.4 dB if we want to be sure that the task was not done monaurally.

Appendix C

Monaural performance and roving in ILD amplification experiments

In this appendix, the effect of level roving to reduce the use of monaural level cues in experiment 2 of chapter 5 is analyzed.

C.1 Introduction

If the loudness of a signal is known, as would be the case in our test setup without roving, the subject could use the head shadow effect monaurally to localize the sound source. While monaural loudness cues are relevant for real-life localization, their salience needs to be reduced in our test setup because (1) in real-life the loudness of the signal is in most cases not exactly known and (2) we are in the current study investigating interaural level cues. Therefore we introduce level roving.

Calculation of the minimal RMS error that could be obtained using only monaural loudness cues for a certain amount of roving is complicated. However, considering the case with only two loudspeakers, for A_-, the minimal level of the vocoder-ear, A_+ the maximal level and a uniform rove of $\pm R$ (all in dB), the chance of answering correctly is

$$P_{correct} = 1 - P_{wrong} \tag{C.1}$$

$$= 1 - \frac{(A_- + R) - (A_+ - R)}{2R} \tag{C.2}$$

$$\Downarrow$$

$$R = \frac{A_- - A_+}{2P_{wrong} - 2} \tag{C.3}$$

For $P_{wrong} = 0.5$, $A_+ = 6.4$ and $A_- = -8.9$ (these values can be obtained from figure 5.9 in the left panel, for the noise14000 signal, in the

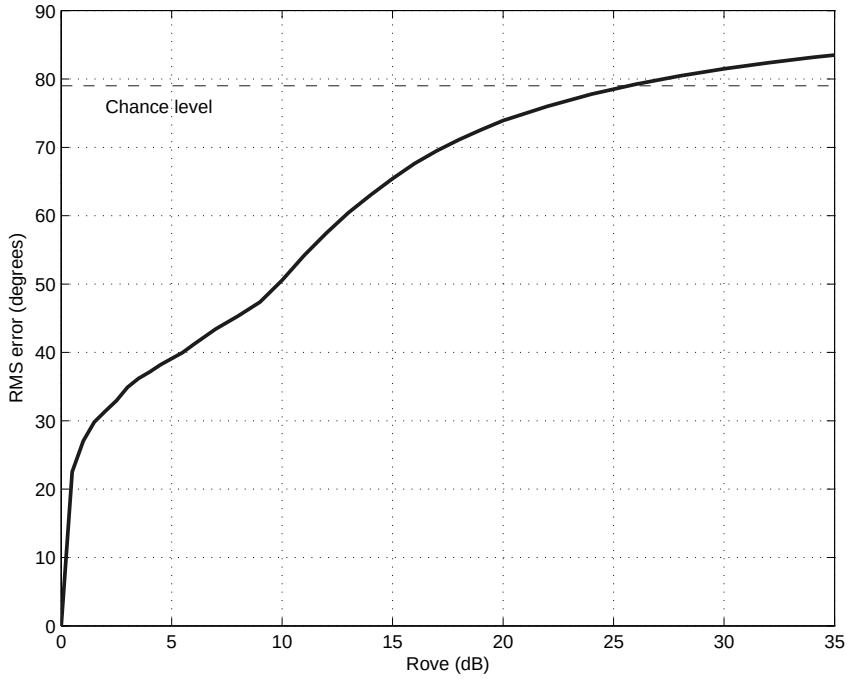


Figure C.1: Monte Carlo simulations of average RMS error obtained with a decision strategy only using monaural loudness cues for different roving ranges (R). Each data point is the median of 10^5 simulations.

"vocoder L" condition), this yields R = 15.3 dB, which would give a total roving range of 2R = 30.6 dB.

The case with 13 loudspeakers was simulated using Monte Carlo simulations. The computer used a list of monaural levels per angle. The decision strategy was for a given monaural level to select the angle closest in level from the list. The list of monaural levels is what a subject could have learned after training. For different roving ranges R, the median RMS error of 10^5 simulations was calculated and is shown in figure C.1. It is clear that the RMS error that can be obtained with this decision strategy increases non-linearly with increasing roving range and that a roving range of R = 25 dB is necessary to degrade performance to chance level.

Such large roving ranges are not feasible in the current study because

Condition	ILDs	Monaural head shadow effect cues	Monaural spectral cues
binaural-amp	*	*	*
binaural-noamp	*	*	*
monaural		*	*
monaural-NLD			*

Table C.1: Localization cues available in the different conditions.

performance decreases with increasing roving level (Francart and Wouters, 2007) and audibility and uncomfortable loudnesses become problematic. Therefore in the current study, smaller roving ranges were used. For experiment 2 a roving range of R = 6 dB was used. In this appendix we assess its influence on our results. The RMS error for R = 6 dB from the latter simulations is 41°.

C.2 Methods

To assess the influence of monaural head shadow cues on our results, experiment 2 was repeated monaurally (using only the vocoder ear) with 5 subjects in a condition with and without monaural head shadow cues between stimuli. The conditions binaural-amp and binaural-noamp correspond to experiment 2. Condition monaural is the same as binaural-noamp, but without the low pass filtered ear, i.e., only the vocoder ear. Condition monaural-NLD is the same as binaural-noamp, but without monaural level differences stemming from the head shadow effect. Roving of ±6 dB was done in all conditions. An overview of the different conditions and available localization cues is given in table C.1.

C.3 Results and discussion

In figure C.2 the results are shown. Conditions binaural-amp and binaural-noamp are respectively the results from experiment 2 with and without ILD amplification.

For the wideband noise signal (noise14000), comparing conditions binaural-noamp and monaural, performance decreased only slightly when removing

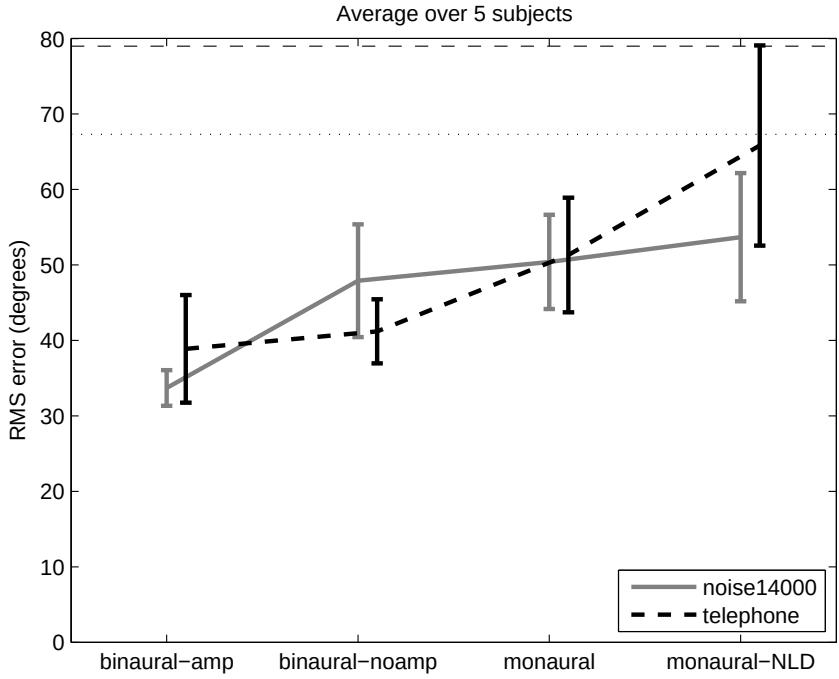


Figure C.2: Comparison of binaural and monaural results. The dotted and dashed lines show the significance and chance level, respectively.

the low pass filtered ear. This means that performance in both conditions was probably largely based on monaural cues (both spectral and level). Comparing conditions monaural and monaural-NLD, indicates that removal of monaural level cues still decreased performance only slightly. While the roving might not have eliminated all monaural level cues, these cues could not improve performance more than monaural spectral cues could. An ANOVA and Tukey post-hoc tests with factors condition and subject show significant differences between binaural-noamp and monaural-NLD ($F(2, 31) = 3.61, p = 0.04$).

For the telephone signal, the result is different. Comparing conditions binaural-noamp and monaural, performance is seen to decrease with the removal of the low pass filtered ear. This is probably caused by less clear monaural spectral cues because the signal has a narrower and less flat spectrum. When removing monaural level cues, performance decreases further. This indicates that monaural level cues were used to achieve the performance in the conditions binaural-noamp and monaural. Van Wanrooij and Van Opstal (2004) show the same reliance on monaural head shadow cues in monaurally deaf subjects. An ANOVA and Tukey post-hoc tests with factors condition and subject indicates significant differences between all conditions ($F(2, 27) = 53.30, p < 0.001$).

C.4 Conclusions

Both monaural level cues stemming from the head shadow effect and monaural spectral cues play a role in localization through the noise band vocoder. As the same amount of roving was used in the conditions with and without ILD amplification of experiment 2, the observed differences in localization performance are valid, but could further increase if monaural level cues were completely eliminated, especially for the telephone signal.

Bibliography

P.J. Abbas and C.J. Brown. Electrically evoked auditory brainstem response: growth of response with current level. *Hear Res*, 51(1):123–37, 1991.

M.A. Akeroyd. The psychoacoustics of binaural hearing. *Int J Audiol*, 45 Suppl 1:25–33, 2006.

N. Bauman. *Ototoxic Drugs Exposed: Prescription Drugs and Other Chemicals That Can (and Do) Damage Our Ears*. GuidePost Publications, 2004.

U. Baumann and A. Nobbe. The cochlear implant electrode-pitch function. *Hear Res*, 213(1-2):34–42, 2006.

L.R. Bernstein and C. Trahiotis. Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli". *J Acoust Soc Am*, 112(3 Pt 1):1026–36, 2002.

L.R. Bernstein and C. Trahiotis. The apparent immunity of high-frequency "transposed" stimuli to low-frequency binaural interference. *J Acoust Soc Am*, 116(5):3062–9, 2004.

L.R. Bernstein and C. Trahiotis. Measures of extents of laterality for high-frequency "transposed" stimuli under conditions of binaural interference. *J Acoust Soc Am*, 118(3 Pt 1):1626–35, 2005.

L.R. Bernstein and C. Trahiotis. Lateralization of low-frequency, complex waveforms: the use of envelope-based temporal disparities. *J Acoust Soc Am*, 77(5):1868–80, 1985a.

L.R. Bernstein and C. Trahiotis. Lateralization of sinusoidally amplitude-modulated tones: effects of spectral locus and temporal variation. *J Acoust Soc Am*, 78(2):514–23, 1985b.

L.R. Bernstein and C. Trahiotis. Why do transposed stimuli enhance binaural processing?: Interaural envelope correlation vs envelope normalized fourth moment. *J Acoust Soc Am*, 121(1):EL23–EL28, 2007.

P. Bertelson. *Cognitive contributions to the perception of spatial and temporal events*. Elsevier Science, 1999.

P.J. Blamey, G.J. Dooley, E.S. Parisi, and G.M. Clark. Pitch comparisons of acoustically and electrically evoked auditory sensations. *Hear Res*, 99(1-2):139–50, 1996.

D.A. Blanks, J.M. Roberts, E. Buss, J.W. Hall, and D.C. Fitzpatrick. Neural and behavioral sensitivity to interaural time differences using amplitude modulated tones with mismatched carrier frequencies. *J Assoc Res Otolaryngol*, 8(3):393–408, 2007.

J. Blauert. *Spatial Hearing*. MIT Press, 1997.

C. Boex, L. Baud, G. Cosendai, A. Sigrist, M.I. Kos, and M. Pelizzone. Acoustic to Electric Pitch Comparisons in Cochlear Implant Subjects with Residual Hearing. *J Assoc Res Otolaryngol*, 7(2):110–24, 2006.

J. Breebaart, S. van de Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J Acoust Soc Am*, 110(2):1089–104, 2001.

A. W. Bronkhorst. Localization of real and virtual sound sources. *J Acoust Soc Am*, 98(5):2542–2553, 1995.

D.S. Brungart. Auditory localization of nearby sources. III. Stimulus effects. *J Acoust Soc Am*, 106(6):3589–602, 1999.

D.S. Brungart and W.M. Rabinowitz. Auditory localization of nearby sources. Head-related transfer functions. *J Acoust Soc Am*, 106(3 Pt 1): 1465–79, 1999.

D.S. Brungart, N.I. Durlach, and W. M. Rabinowitz. Auditory localization of nearby sources. II. Localization of a broadband source. *J Acoust Soc Am*, 106(4 Pt 1):1956–68, 1999.

S. Buus. Level discrimination of frozen and random noise. *J Acoust Soc Am*, 87(6):2643–54, 1990.

M. Chatterjee, Q.J. Fu, and R.V. Shannon. Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners. *J Acoust Soc Am*, 107(3):1637–44, 2000.

T.Y. Ching, C. Psarros, M. Hill, H. Dillon, and P. Incerti. Should children who use cochlear implants wear hearing aids in the opposite ear? *Ear Hear*, 22(5):365–80, 2001.

T.Y. Ching, P. Incerti, and M. Hill. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear*, 25(1): 9–21, 2004.

T.Y. Ching, E. van Wanrooy, and H. Dillon. Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: a review. *Trends Amplif*, 11(3):161–92, 2007.

G. Clark. *Cochlear Implants: Fundamentals and Applications*. Springer, 2003.

L.T. Cohen, J. Xu, S.A. Xu, and G.M. Clark. Improved and simplified methods for specifying positions of the electrode bands of a cochlear implant array. *Am J Otol*, 17(6):859–65, 1996.

S. Colburn, B. Shinn-Cunningham, G.J. Kidd, and N. Durlach. The perceptual consequences of binaural hearing. *Int J Audiol*, 45 Suppl 1: 34–44, 2006.

T.H. Cormen, C.E. Leiserson, R L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press & McGraw-Hill, 2nd edition, 2001.

H. Dillon. *Hearing Aids*. Thieme, New York, 2001.

R. Domnitz. The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. *J Acoust Soc Am*, 53(6):1549–52, 1973.

M. Don and J.J. Eggermont. Analysis of the click-evoked brainstem potentials in man using high-pass noise masking. *J Acoust Soc Am*, 63(4): 1084–92, 1978.

M.F. Dorman, L. Smith, and J.L. Parkin. Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant. *Ear Hear*, 14(4):290–2, 1993.

M.F. Dorman, M. Smith, L. Smith, and J.L. Parkin. The pitch of electrically presented sinusoids. *J Acoust Soc Am*, 95(3):1677–9, 1994.

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. *J Acoust Soc Am*, 102(4): 2403–2411, 1997.

M.F. Dorman, R.H. Gifford, A.J. Spahr, and S.A. McKarns. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurootol*, 13(2):105–112, 2007a.

M.F. Dorman, T. Spahr, R. Gifford, L. Loiselle, S. McKarns, T. Holden, M. Skinner, and C. Finley. An Electric Frequency-to-place Map for a Cochlear Implant Patient with Hearing in the Nonimplanted Ear. *J Assoc Res Otolaryngol*, 8(2):234–40, 2007b.

C.C. Dunn, R.S. Tyler, and S.A. Witt. Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant. *J Speech Lang Hear Res*, 48(3):668–80, 2005.

D. Eddington, J. Tierney, V. Noel, B. Herrmann, M. Whearty, and C.C. Finley. Speech processors for auditory prostheses. *NIH quarterly progress report*, Contract N01-DC-(3), 2002.

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. *Ann Otol Rhinol Laryngol*, 87(6 Pt 2):1–39, 1978.

W.E. Feddersen, T.T. Sandel, D.C. Teas, and L. A. Jeffress. Localization of high-frequency tones. *J Acoust Soc Am*, 29(9):988–991, 1957.

B.J. Fligor and L.C. Cox. Output levels of commercially available portable compact disc players and the potential risk to hearing. *Ear Hear*, 25(6): 513–27, 2004.

T. Francart and J. Wouters. Perception of across-frequency interaural level differences. *J. Acoust. Soc. Am.*, 122(5):2826–2831, 2007.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. *Audiol Neurootol.*, 13(5):309–319, 2008a.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation. *J Assoc Res Otolaryngol*, In press, 2008b.

T. Francart, M. Moonen, and J. Wouters. Automatic testing of speech recognition. *Int J Audiol*, In press, 2008c.

T. Francart, T. Van den Bogaert, M. Moonen, and J. Wouters. Amplification of interaural level differences improves sound localization for cochlear implant users with contralateral acoustic hearing. *J Acoust Soc Am*, conditionally accepted, 2008d.

T. Francart, A. van Wieringen, and J. Wouters. APEX 3: a multi-purpose test platform for auditory psychophysical experiments. *J Neurosci Methods*, 172(2):283–293, 2008e.

J.E. F. Friedl. *Mastering Regular Expressions*. O'Reilly, 3rd edition, 2006.

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. *J Acoust Soc Am*, 110(2): 1150–1163, 2001.

Q.J. Fu. Loudness growth in cochlear implants: effect of stimulation rate and electrode configuration. *Hear Res*, 202(1-2):55–62, 2005.

Q.J. Fu and R.V. Shannon. Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. *J Acoust Soc Am*, 104(5):2570–7, 1998.

Q.J. Fu, R.V. Shannon, and J.J. Galvin. Perceptual learning following changes in the frequency-to-electrode assignment with the nucleus-22 cochlear implant. *J Acoust Soc Am*, 112(4):1664–74, 2002.

S. Gallego, S. Garnier, C. Micheyl, E. Truy, A. Morgon, and L. Collet. Loudness growth functions and EABR characteristics in Digisonic cochlear implantees. *Acta Otolaryngol*, 119(2):234–8, 1999.

B.J. Gantz and C. Turner. Combining acoustic and electrical speech processing: Iowa/Nucleus hybrid implant. *Acta Otolaryngol*, 124(4):344–7, 2004.

L. Geurts and J. Wouters. A concept for a research tool for experiments with cochlear implant users. *J Acoust Soc Am*, 108(6):2949–56, 2000.

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

D.W. Grantham, D.H. Ashmead, T.A. Ricketts, R.F. Labadie, and D.S. Haynes. Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants. *Ear Hear*, 28(4):524–41, 2007a.

D.W. Grantham, T.A. Ricketts, D.H. Ashmead, R.F. Labadie, and D.S. Haynes. Localization by postlingually deafened adults fitted with a single cochlear implant. *Laryngoscope*, 2007b.

D.D. Greenwood. A cochlear frequency-position function for several species–29 years later. *J Acoust Soc Am*, 87(6):2592–605, 1990.

W. Gstoettner, J. Kiefer, W.D. Baumgartner, S. Pok, S. Peters, and O. Adunka. Hearing preservation in cochlear implantation for electric acoustic stimulation. *Acta Otolaryngol*, 124(4):348–52, 2004.

W.K. Gstoettner, S. Helbig, N. Maier, J. Kiefer, A. Radeloff, and O.F. Adunka. Ipsilateral electric acoustic stimulation of the auditory system: results of long-term hearing preservation. *Audiol Neurootol*, 11 Suppl 1: 49–56, 2006.

E.R. Hafter and S.C. Carrier. Binaural interaction in low-frequency stimuli: the inability to trade time and intensity completely. *J Acoust Soc Am*, 51(6):1852–62, 1972.

E.R. Hafter, R.H. Dye, Jr., E.M. Wenzel, and K. Knecht. The combination of interaural time and intensity in the lateralization of high-frequency complex signals. *J Acoust Soc Am*, 87(4):1702–8, 1990.

W.M. Hartmann. How we localize sound. *Physics Today*, 11:24–29, 1999.

W.M. Hartmann and Z.A. Constan. Interaural level differences and the level-meter model. *J Acoust Soc Am*, 112(3 Pt 1):1037–45, 2002.

W.M. Hartmann and B. Rakerd. On the minimum audible angle–a decision theory approach. *J Acoust Soc Am*, 85(5):2031–41, 1989.

G.B. Henning. Detectability of interaural delay in high-frequency complex waveforms. *J Acoust Soc Am*, 55(1):84–90, 1974.

P.M. Hofman, R.J. Van, and O.A. Van. Relearning sound localization with new ears. *Nat Neurosci*, 1(5):417–21, 1998.

I. Holube, M. Kinkel, and B. Kollmeier. Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments. *J Acoust Soc Am*, 104(4):2412–25, 1998.

S. Hoth. Indication for the need of flexible and frequency specific mapping functions in cochlear implant speech processors. *Eur Arch Otorhinolaryngol*, 264(2):129–38, 2007.

C. James, P. Blamey, J.K. Shallop, P.V. Incerti, and A.M. Nicholas. Contralateral masking in cochlear implant users with residual hearing in the non-implanted ear. *Audiol Neurootol*, 6(2):87–97, 2001.

A.R. Javer and D.W. Schwarz. Plasticity in human directional hearing. *J Otolaryngol*, 24:111–117, 1995.

J. Kiefer, M. Pok, O. Adunka, E. Sturzebecher, W. Baumgartner, M. Schmidt, J. Tillein, Q. Ye, and W. Gstoettner. Combined Electric and Acoustic Stimulation of the Auditory System: Results of a Clinical Study. *Audiol Neurootol*, 10(3):134–144, 2005.

Y.Y. Kong and R. P. Carlyon. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *J Acoust Soc Am*, 121(6):3717–27, 2007.

Y.Y. Kong, G.S. Stickney, and F.G. Zeng. Speech and melody recognition in binaurally combined acoustic and electric hearing. *J Acoust Soc Am*, 117(3 Pt 1):1351–61, 2005.

R. Krahe, O.N. Larsen, and B. Ronacher. Directional hearing is only weakly dependent on the rise time of acoustic stimuli. *J Acoust Soc Am*, 107(2):1067–70, 2000.

G. F. Kuhn. Model for the interaural time differences in the azimuthal plane. *J Acoust Soc Am*, 62(1):157–167, 1977.

K. Kukich. Technique for automatically correcting words in text. *ACM Comput. Surv.*, 24(4):377–439, 1992. ISSN 0360-0300.

B. Laback, S.M. Pok, W.D. Baumgartner, W.A. Deutsch, and K. Schmid. Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors. *Ear Hear*, 25(5):488–500, 2004.

B. Laback, P. Majdak, and W.D. Baumgartner. Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing. *J Acoust Soc Am*, 121(4):2182–91, 2007.

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. *J Neurosci Meth*, 142(1):131–6, 2005.

E.H. Langendijk and A.W. Bronkhorst. Contribution of spectral cues to human sound localization. *J Acoust Soc Am*, 112(4):1583–96, 2002.

D.T. Lawson, B.S. Wilson, M. Zerbi, C. van den Honert, C.C. Finley, J.C. Farmer, Jr, J.T. McElveen, Jr, and P.A. Roush. Bilateral cochlear implants controlled by a single speech processor. *Am J Otol*, 19(6): 758–61, 1998.

M.R. Leek. Adaptive procedures in psychophysical research. *Percept Psychophys*, 63(8):1279–92, 2001.

E.L. LePage and N.M. Murray. Latent cochlear damage in personal stereo users: a study based on click-evoked otoacoustic emissions. *Med. J. Aust.*, 169:588–592, 1998.

V. Levenshtein. Binary codes capable of correcting spurious insertions and deletions or ones. *Problems of information transmission*, (1):8–17, 1965.

H. Levitt. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, 49(2):467–477, 1971.

J. Lewald and S. Getzmann. Horizontal and vertical effects of eye-position on sound localization. *Hear Res*, 213(1-2):99–106, 2006.

R. Litovsky, P.M. Johnstone, and S.P. Godar. Benefits of bilateral cochlear implants and/or hearing aids in children. *Int J Audiol*, 45 Suppl 1:78–91, 2006.

C. Long, R. P. Carlyon, R. Litovsky, and D. Downs. Binaural unmasking with bilateral cochlear implants. *J Assoc Res Otolaryngol*, 7(4):352–60, 2006.

C.J. Long, D.K. Eddington, H.S. Colburn, and W.M. Rabinowitz. Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. *J Acoust Soc Am*, 114(3):1565–74, 2003.

E.A. Macpherson and J.C. Middlebrooks. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am*, 111(5 Pt 1):2219–36, 2002.

P. Majdak, B. Laback, and W.D. Baumgartner. Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing. *J Acoust Soc Am*, 120(4):2190–201, 2006.

H.J. McDermott and C.M. Sucher. Perceptual dissimilarities among acoustic stimuli and ipsilateral electric stimuli. *Hear Res*, 218(1-2):81–8, 2006.

D. McFadden and E.G. Pasanen. Lateralization of high frequencies based on interaural time differences. *J Acoust Soc Am*, 59(3):634–9, 1976.

H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.

J.C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am*, 106(3 Pt 1):1493–510, 1999.

A.W. Mills. On the minimum audible angle. *J Acoust Soc Am*, 30(237), 1958.

A.W. Mills. Lateralization of high-frequency tones. *J Acoust Soc Am*, 32 (132):132–134, 1960.

P. Minnaar, S.K. Olesen, F. Christensen, and H. Møller. Localization with binaural recordings from artificial and human heads. *Journal of the Audio Engineering Society*, 49(5):323–336, 2001.

B.C.J. Moore. *Perceptual Consequences of Cochlear Damage.* Oxford Medical Publications, 1995.

B.C.J. Moore. *An introduction to the Psychology of Hearing, 5th edition.* Elsevier Science, 2003.

A.D. Musicant and R.A. Butler. The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am*, 75(4):1195–200, 1984.

G. Navarro. A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1):31–88, 2001.

G.C. Nikiforidis, C.M. Koutsojannis, J.N. Varakis, and P.D. Goumas. Reduced variance in the latency and amplitude of the fifth wave of auditory brain stem response after normalization for head size. *Ear Hear*, 14(6): 423–8, 1993.

H. Noh, P.J. Abbas, C.A. Abbas, K.V. Nourski, B.K. Robinson, and F.C. Jeng. Binaural interactions of electrically and acoustically evoked responses recorded from the inferior colliculus of guinea pigs. *Int J Audiol*, 46(6):309–20, 2007.

P. Nopp, P. Schleich, and P. D'Haese. Sound localization in bilateral users of med-el combi 40/40 cochlear implants. *Ear Hear*, 25(3):205–14, 2004.

J.M. Nuetzel and E.R. Hafter. Lateralization of complex waveforms: effects of fine structure, amplitude, and duration. *J Acoust Soc Am*, 60 (6):1339–46, 1976.

J.M. Nuetzel and E.R. Hafter. Discrimination of interaural delays in complex waveforms: Spectral effects. *J Acoust Soc Am*, 69(4):1112–1118, 1981.

A.R. Palmer, L.F. Liu, and T.M. Shackleton. Changes in interaural time sensitivity with interaural level differences in the inferior colliculus. *Hear Res*, 223(1-2):105–13, 2007.

R.D. Patterson, M.H. Allerhand, and C. Giguere. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J Acoust Soc Am*, 98(4):1890–4, 1995.

S. Perrett and W. Noble. Available response choices affect localization of sound. *Percept Psychophys*, 57(2):150–8, 1995.

D.P. Phillips and S.E. Hall. Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level. *Hear Res*, 202(1-2):188–99, 2005.

D.P. Phillips, M.E. Carmichael, and S.E. Hall. Interaction in the perceptual processing of interaural time and level differences. *Hear Res*, 211 (1-2):96–102, 2006.

L. Phillips. The Double Metaphone Search Algorithm. *C/C++ Users Journal*, 2000.

G. Plenge. On the differences between localization and lateralization. *J Acoust Soc Am*, 56(3):944–51, 1974.

L.A. Reiss, C.W. Turner, S.R. Erenberg, and B.J. Gantz. Changes in Pitch with a Cochlear Implant Over Time. *J Assoc Res Otolaryngol*, 8 (2):241–257, 2007.

M. Reynaert. Multilingual text induced spelling correction. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, 2004.

M. Reynaert. *Text-induced spelling correction*. PhD thesis, University of Tilburg, 2005.

S. Rosen, A. Faulkner, and L. Wilkinson. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am*, 106(6):3629–36, 1999.

D. Rowan and M.E. Lutman. Learning to discriminate interaural time differences at low and high frequencies. *Int J Audiol*, 46(10):585–94, 2007.

K. Saberi. Modeling interaural-delay sensitivity to frequency modulation at high frequencies. *J Acoust Soc Am*, 103(5 Pt 1):2551–64, 1998.

B.U. Seeber, U. Baumann, and H. Fastl. Localization ability with bimodal hearing aids and bilateral cochlear implants. *J Acoust Soc Am*, 116(3): 1698–709, 2004.

P. Senn, M. Kompis, M. Vischer, and R. Haeusler. Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors. *Audiol Neurootol*, 10(6):342–52, 2005.

J.K. Shallop, A.L. Beiter, D.W. Goin, and R.E. Mischke. Electrically evoked auditory brain stem responses (EABR) and middle latency responses (EMLR) obtained from patients with the nucleus multichannel cochlear implant. *Ear Hear*, 11(1):5–15, 1990.

R.V. Shannon. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hear Res*, 11(2):157–89, 1983.

R.V. Shannon, F.G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. *Science*, 270:303–304, 1995.

N.T. Shepard and S. Colburn. Interaural time discrimination of clicks: Dependence on interaural time and intensity differences. *J Acoust Soc Am*, 59(S1):S23, 1976.

M. Slaney. An efficient implementation of the Patterson-Holdsworth auditory filterbank. *Apple Tech Rep*, (35), 1993.

C. Smits, P. Merkus, and T. Houtgast. How we do it: The dutch functional hearing-screening tests by telephone and internet. *Clin Otolaryngol*, 31 (5):436–40, 2006.

G.S. Stickney, F.G. Zeng, R. Litovsky, and P. Assmann. Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am*, 116(2):1081–91, 2004.

G.S. Stickney, K. Nie, and F.G. Zeng. Contribution of frequency modulation to speech recognition in noise. *J Acoust Soc Am*, 118(4):2412–20, 2005.

J.W. Strutt. *The theory of sound*. Dover Publications, 1877.

C.W. Turner, B.J. Gantz, C. Vidal, A. Behrens, and B.A. Henry. Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. *J Acoust Soc Am*, 115(4):1729–35, 2004.

R.S. Tyler, A.J. Parkinson, B.S. Wilson, S. Witt, J.P. Preece, and W. Noble. Patients utilizing a hearing aid and a cochlear implant: speech perception and localization. *Ear Hear*, 23(2):98–105, 2002.

T. Van den Bogaert, T. Klasen, M. Moonen, L. Van Deun, and J. Wouters. Horizontal localisation with bilateral hearing aids: without is better than with. *J Acoust Soc Am*, 119(1):515–526, 2006.

L. Van Deun, A. van Wieringen, T. Francart, F. Scherf, I. Dhooge, N. Deggouj, C. Desloovere, P. Van de Heyning, F.E. Offeciers, L. De Raeve, and J. Wouters. Bilateral cochlear implants in children: binaural unmasking. *Audiol Neurotol*, Accepted, 2008.

R.J. van Hoesel. Exploring the benefits of bilateral cochlear implants. *Audiol Neurootol*, 9(4):234–46, 2004.

R.J. van Hoesel. Sensitivity to binaural timing in bilateral cochlear implant users. *J Acoust Soc Am*, 121(4):2192–206, 2007.

R.J. van Hoesel and R.S. Tyler. Speech perception, localization, and lateralization with bilateral cochlear implants. *J Acoust Soc Am*, 113(3): 1617–30, 2003.

R.J. Van Hoesel, R. Ramsden, and M. Odriscoll. Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear Hear*, 23 (2):137–49, 2002.

M. Van Wanrooij and A. Van Opstal. Contribution of head shadow and pinna cues to chronic monaural sound localization. *J Neurosci*, 24(17): 4163–71, 2004.

A. van Wieringen and J. Wouters. LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and The Netherlands. *Int J Audol*, 47(6):348–355, 2008.

C.A. Verschuur, M.E. Lutman, R. Ramsden, P. Greenham, and O. M. Auditory localization abilities in bilateral cochlear implant recipients. *Otol Neurotol*, 26(5):965–71, 2005.

N.J. Versfeld, L. Daalder, J.M. Festen, and T. Houtgast. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, 107(3):1671–84, 2000.

K.C. Wagener, T. Brand, and B. Kollmeier. The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *Int J Audiol*, 45(1):26–33, 2006.

H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *J of Exp Psychol*, 27(4):339–368, 1940.

E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman. Localization using nonindividualized head-related transfer functions. *J Acoust Soc Am*, 94(1):111–23, 1993.

F.A. Wichmann and N.J. Hill. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63(8):1314–1329, 2001.

F.L. Wightman and D.J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J Acoust Soc Am*, 105 (5):2841–53, 1999.

F.L. Wightman and D.J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J Acoust Soc Am*, 91 (3):1648–61, 1992.

F.L. Wightman and D.J. Kistler. Headphone simulation of free-field listening. i: Stimulus synthesis. *J Acoust Soc Am*, 85(2):858–67, 1989a.

F.L. Wightman and D.J. Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. *J Acoust Soc Am*, 85(2):868–78, 1989b.

W. Williams. Noise exposure levels from personal stereo use. *Int J Audiol*, 44(4):231–6, 2005.

J. Wouters, W. Damman, and A.J. Bosman. Vlaamse opname van woordenlijsten voor spraakaudiometrie. *Logopedie*, (7):28–33, 1994.

B.A. Wright and M.B. Fitzgerald. Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proc Natl Acad Sci U S A*, 98(21):12307–12, 2001.

B.A. Wright and Y. Zhang. A review of learning with normal and altered sound-localization cues in human adults. *Int J Audiol*, 45 Suppl 1:92–8, 2006.

W. A. Yost. Lateral position of sinusoids presented with interaural intensive and temporal differences. *J Acoust Soc Am*, 70(2):397–409, 1981.

W.A. Yost. Discriminations of interaural phase differences. *J Acoust Soc Am*, 55(6):1299–303, 1974.

W.A. Yost and R.H. Dye, Jr. Discrimination of interaural differences of level as a function of frequency. *J Acoust Soc Am*, 83(5):1846–51, 1988.

W.A. Yost, F.L. Wightman, and David M. Green. Lateralization of filtered clicks. *J Acoust Soc Am*, 50(6B):1526–1531, 1971.

P. Zahorik, P. Bangayan, V. Sundareswaran, K. Wang, and C. Tam. Perceptual recalibration in human sound localization: learning to remediate front-back reversals. *J Acoust Soc Am*, 120(1):343–59, 2006.

F.G. Zeng and R.V. Shannon. Loudness balance between electric and acoustic stimulation. *Hear Res*, 60(2):231–5, 1992.

F.G. Zeng and R.V. Shannon. Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. *Science*, 264(5158): 564–6, 1994.

F.G. Zeng, A.N. Popper, and R.R. Fay. *Cochlear implants*. Springer, 2004.

J. Zwislocki and R.S. Feldman. Just noticeable differences in dichotic phase. *J Acoust Soc Am*, 28(5):860–864, 1956.

List of publications

Publications in international journals

T. Francart and J. Wouters. Perception of across-frequency interaural level differences. *J. Acoust. Soc. Am.*, 122(5):2826–2831, 2007.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. *Audiol Neurootol.*, 13(5):309–319, 2008a.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation. *J Assoc Res Otolaryngol*, In press, 2008b.

T. Francart, M. Moonen, and J. Wouters. Automatic testing of speech recognition. *Int J Audiol*, In press, 2008c.

T. Francart, T. Van den Bogaert, M. Moonen, and J. Wouters. Amplification of interaural level differences improves sound localization for cochlear implant users with contralateral acoustic hearing. *J Acoust Soc Am*, conditionally accepted, 2008d.

T. Francart, A. van Wieringen, and J. Wouters. APEX 3: a multi-purpose test platform for auditory psychophysical experiments. *J Neurosci Methods*, 172(2):283–293, 2008e.

L. Van Deun, A. van Wieringen, T. Francart, F. Scherf, I. Dhooge, N. Deggouj, C. Desloovere, P. Van de Heyning, F.E. Offeciers, L. De Raeve, and J. Wouters. Bilateral cochlear implants in children: binaural unmasking. *Audiol Neurotol*, Accepted, 2008.

Abstracts in conference proceedings

K. Eneman and T. Francart. Analyse van de zangstem: geluiden in beeld. In F. de Jong and W. Decoster, editors, *STEM*, Leuven, Belgium, 2008.

T. Francart and K. Eneman. Analyse van de zangstem: geluiden in beeld. In *Bridging voice professionals & VOX 2007*, Leuven, Belgium, 2007.

T. Francart and J. Wouters. Noise band vocoder simulations of electric acoustic stimulation. In *Conference on Implantable Auditory Prosthesis*, Asilomar, California, USA, 2005.

T. Francart and J. Wouters. Perception of across-frequency interaural level differences. In *International Hearing Aid Research Conference*, Lake Tahoe, California, USA, 2006.

T. Francart and J. Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. In *Conference on Implantable Auditory Prostheses*, Lake Tahoe, California, USA, 2007.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural time differences with bilateral bimodal stimulation. In *International Hearing Aid Research Conference*, Lake Tahoe, California, USA, 2008a.

T. Francart, M. Moonen, and J. Wouters. Automatic testing of speech understanding. *J Acoust Soc Am*, 123(5):3065–3065, 2008b.

Conference posters

T. Francart and J. Wouters. Noise band vocoder simulations of electric acoustic stimulation. Poster, Conference on Implantable Auditory Prosthesis, Asilomar, California, USA, 2005.

T. Francart and J. Wouters. Perception of across-frequency interaural level differences. Poster, International Hearing Aid Research Conference, Lake Tahoe, California, USA, 2006a.

T. Francart and J. Wouters. Horen met een cochleair implantaat en hoorapparaat samen. Poster, Symposium Logopedische en Audiologische wetenschappen, Leuven, Belgium, 2006b.

T. Francart, J. Brokx, and J. Wouters. Sensitivity to interaural time differences with bilateral bimodal stimulation. Poster, International Hearing Aid Research Conference, Lake Tahoe, California, USA, 2008.

Conference presentations

K. Eneman and T. Francart. Analyse van de zangstem: geluiden in beeld. Presentation, Stem symposium, Leuven, Belgium, 2008.

T. Francart and K. Eneman. Analyse van de zangstem: geluiden in beeld. Presentation, VOX symposium, Leuven, Belgium, 2007.

T. Francart and J. Wouters. Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation. Presentation, Conference on Implantable Auditory Prostheses, Lake Tahoe, California, 2007.

T. Francart and J. Wouters. APEX 3 and NICv2. Presentation, NIC workshop, Mechelen, Belgium, 2005.

T. Francart and J. Wouters. Localization with bimodal stimulation. Presentation, NIC workshop, Mechelen, Belgium, 2006.

T. Francart and J. Wouters. Perception of binaural cues with bimodal hearing (CI+HA). Presentation, WAS dag, UMC, Utrecht, Netherlands, 2008.

T. Francart, M. Moonen, and J. Wouters. Automatic testing of speech understanding. Presentation, Acoustics, Paris, France, 2008.

J. Wouters and T. Francart. Presentation, 30 Jaar CI in België, Brussels, Belgium, 2005.

Curriculum Vitae

Tom Francart was born in Leuven, Belgium on 9 April 1981. He lived in Heverlee ever since. In 2004 he wrote his master's thesis on synthesis and analysis of the singing voice and received the degree of Master in Electrical Engineering at the K.U.Leuven.

In the same year, he started his PhD at ExpORL under the supervision of Prof. dr. Jan Wouters and Prof. dr. ir. Marc Moonen. His main research interests are hearing and sound in general and more specifically cochlear implants, hearing aids, sound source localization, speech perception and music perception. He has additional scientific interests in the singing voice and informatics. His personal interests include classical singing and gastronomy.