KU LEUVEN

Applied Physiological Modelling of Auditory Processes

Speech Intelligibility, Modulation Detection, and Binaural Hearing

Arturo Moncada-Torres

November 2018

Applied Physiological Modelling of Auditory Processes

Speech Intelligibility, Modulation Detection, and Binaural Hearing

Arturo Moncada-Torres

Examination committee:
Prof. Dr. Astrid van Wieringen, chair
Prof. Dr. Tom. Francart, supervisor
Prof. Dr. Jan Wouters, co-supervisor
Prof. Dr. Marc Moonen
Prof. Dr. Christian Desloovere
Prof. Dr. Torsten Dau
(Technical University of Denmark)
Prof. Dr. Sarah Verhulst
(Ghent University)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Science (PhD): Biomedical Sciences

November 2018

© 2018 KU Leuven – Faculty of Medicine
Self-published, Arturo Moncada-Torres, Herestraat 49, Bus 721, B-3000 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, 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, electronic or any other means without written permission from the publisher.

Acknowledgements

Well, this is it. This thesis is the product of the a-little-bit-over 4 years that I spent as a PhD student in Leuven, Belgium. Scientifically, it was a very challenging and incredibly rewarding endeavour. Most importantly, personally my time here has been an unforgettable experience. All of this wouldn't have been possible without the support of so many people to whom I would like to give proper thanks.

First of all, I would like to thank my (co-)supervisors, Prof. Tom Francart and Prof. Jan Wouters. Tom, thank you very much for your constant support through these years. I appreciate so much being able to sit down with you week after week to discuss the latest results. Furthermore, your patience and empathy were fundamental, specially when things weren't going so well. I consider myself very lucky for having the chance to have a supervisor like you. Jan, thanks for your guidance. Your experience in being able to see the big picture was fundamental to steer my PhD in the right direction.

To all my jury members, Prof. Marc Moonen, Prof. Christian Desloovere, Prof. Torsten Dau, and Prof. Sarah Verhulst, thank you very much for your time and effort in revising this thesis. It is a great honor to have such experts in the field as yourselves as part of my jury committee.

I cannot go any further without thanking one of the finest people I know: Prof. Roger Gassert. Roger, although you weren't directly involved in this thesis, your support has been a cornerstone of my professional career. Thank you so much for believing in this *amigo* and for kickstarting my path as a researcher. I will always appreciate it and never forget it.

My time at ExpORL wouldn't have been half as amazing as it was if I wouldn't have the opportunity to have such wonderful colleagues. First, I would like to start with the group of office nomads. Maaike, although you've been gone for a while, I still appreciate you being my desk buddy for most of my time here. Thanks for trying to make me a healthier person (which kinda worked), for helping me with my (probably stupid) questions about audiology, and for introducing me to the singular world of cantus. Robin, thanks for your help and support during this last stretch of my time here. Most importantly, thank you for our talks about TV shows and football. Going out for dinner and then to catch the latest superhero blockbuster was always something that I enjoyed very much and that I will miss even more. Hanne, you are one of the first friends that I made when I arrived to Belgium. Thanks for making me feel welcome in a new country. Our swingrock lessons were always super fun. Dimitar, thanks for always being willing to quickly answer all my questions, whenever I had one. Anastasiya, Annelore, Charlotte, Damien, Eline, Jana, and Neetha, thanks for being part of the small ExpORL3 family and for making the office a fun space. Most importantly, thanks for bringing all the food. It was always very much appreciated. I will miss all of you. Sanne, thank you very much for all your help when it came to APEX issues. Ania, Annelies, Ben, Benjamin, and Lien, thanks for all the socials and for helping me clear my head when work was just too much. I would like to give a special shout out to the people I had the luck to collaborate with. Andreas, working with you was a blast. Thanks for all your help, support, emails, and Skype calls product of the many iterations of our papers. It was very nice to have someone else in the lab with whom I could talk to about models. What's even nicer is that such a collaboration ended in a good friendship. Sara, thanks for being my paper buddy. I hope that you enjoyed our project together as much as I did. Tilde, I was very lucky to have the chance to work with you. Thanks for all the laughs when trying to have a working protocol. Special thanks to Frieda, Martine, and Olivia for making my life much easier when it came to administrative procedures. To everyone else, thanks for making ExpORL the great lab that it is.

One of the things that I enjoyed the most of my experience as a PhD student was being part of the *ICanHear* project. Not only was it a great platform for training different skills and to share ideas, but it was also a fantastic group of people that made each scientific meeting something to look forward to. Furthermore, this platform gave me the chance to spend six months at the Danish Technical University as part of my secondment, which was amazing. I would like to give special thanks to some people there. Bastian, thanks for your guidance and supervision during my time there. Suyash, I appreciate very much our brainstorming sessions. Thanks for your help, time, and for making me feel included as part of the Hearing Systems group. Christoph, although we didn't work together directly, I value our scientific and non-scientific conversations very much.

During my time in Leuven, Thursday nights were always sacred: volleyball time. Not only did I reconnect with my favorite sport, but I also had the chance to be part of a fantastic group of people. Antoine, Anton, Bertram, Bram, Cyrielle, Frank, Hanne, Inge, Irena, Irma, Jan, Luke, Piet, Sergey, Sigfried, Wim: thank you for letting me have something to really look forward to every week. Our team was where I first felt part of a group after moving to Belgium. A friendly (and exciting) volleyball match plus beers afterwards was always the perfect combo. I will miss you very much.

Furthermore, I also had the opportunity to discover a great hobby in the shape of dancing. Learning how to shake the booty was a great way to blow off some steam and always fun. It also gave me the opportunity to meet some great people. Christian, Kirstine, Laura, Pernille, Søren, Stephanie: *tak* for being so *hygge* (I hope I am using the word right, I meant “warm and friendly”) and for sharing with me my first steps (no pun intended) in Denmark. Coming back to Belgium, thanks to Daisy, Elise, Elke, Heleen, Jasmien, Lore, Manith, Mercedes, and Veronika for the incredibly fun times we had during our classes. Furthermore, thanks to Alex, Antonio, Blanca, Carl, and Gaëtan for their patience when teaching those classes. Having taught me how to dance, even if just a little, is no small feat, I promise.

Thanks to everyone that I've met since I started my new job at IKNL. I appreciate feeling so welcome in a new field, a new group, and a new country. Special thanks to Elsemiek, Gijs, Marjon, Valery, and Xander for their help and for the opportunity to join to such exciting and noble endeavour.

I wouldn't be who I am without the most amazing people that I am proud to call *mis amigos y amigas*. I was lucky enough to cross paths with them and even more to still have them in my life, no matter the distance nor the time zone difference. Waking up to hundreds of text messages blabbing about everything and nothing at the same time made my day every single time. To everyone of *Los Amigos del Árbol* (Carlos, Elías, Emilio, Joaquín, and Patricio): thank you for more than 25 years of friendship (and counting). The number of times that we've laughed until our bellies hurt are just too many. Thanks for being the same fun, playful, and foolish guys I've always known, but also the loyal and supportive friends that you are when I have needed you the most. To my *Iberocuates* (César, Daniel, Klei, Mafer, Marimar, and Marlene): thank you for your friendship, your jokes, and your support in the good and the not-so-good times. So many things have changed since our time in college (new jobs, marriages, relocations, and who knows what might come next). Nevertheless, it is fantastic that every time that we have the opportunity to meet, we can talk as if nothing has changed at all.

Klára, becoming part of each other's life has been the craziest turn of events. Thanks for all of our cooking evenings, scientific conversations, binge-watching sessions, squash matches, running days (not so many of those, fortunately), and amazing trips. I have no words to describe how much I value and appreciate your company, support, and love, specially during the last few months. They've kept me sane and have make me incredibly happy. I cannot wait for whatever is in store for us in the future.

Familia, thank you for supporting me since, well, always, even if this meant me being far from you or if you weren't completely sure of what is it that I do. *Tía* Gude, you have been much more than an aunt. Thank you for your care and affection. I will always be looking forward to our next *Gudesayuno*. Dad, no matter where you are, I hope that you know that you will always be my role model and that I will never forget your life lessons. I can only imagine how happy you'd be with me finally stopping being a "student" and going into "the real world". I am too, I only wish I could share this happiness with you. Mom, *mi Jechu*, thanks for being the best mother ever. I truly value every morning that you woke up early to take us to school, every meal that you prepared, and every time that you told us that we should wash our hands. Mariana, *hermanita*, my little partner in crime of so many adventures since we were kids until now. Thanks for the laughter and for being the sunshine of this family.

Abstract

In order to understand how we process and perceive sound and speech, we need to complement anatophysiological knowledge with psychoacoustical information. However, relating anatomical structures and their physiology to their contribution to auditory perception is a challenging task.

Computational models are valuable tools that allow to bridge the gap between the anatophysiological knowledge of the auditory system and the psychoacoustics of sound and speech. In the last decades, physiologically-based models have increased in popularity and gained importance in the field. These simulate as closely as possible the anatomical components and the physiological processes that occur in the auditory system. Recently, Zilany et al. (2014, 2009) published a model of the auditory periphery up to the auditory nerve (AN) level. Its responses have been validated over a wide range of physiological data from literature.

In this thesis, we used said model as a front end in the development of different frameworks to study speech intelligibility (SI), spectrotemporal modulation detection (STMD), and interaural time difference (ITD) perception.

First, we assessed SI in noise using well-known objective metrics and compared their performance with our framework. Speech can be decomposed into its envelope (i.e., relatively slow variations in amplitude over time) and its temporal fine structure (i.e., rapid oscillations). At a word level, we found that the physiologically-inspired metrics that incorporated envelope information correlated the strongest with behavioral scores. A multivariate linear regression analysis further endorsed their capability for SI prediction. At a phoneme level, we found that phoneme transitions have a larger impact on the envelope component of speech at the AN level.

SI relies on an individual's sensitivity to spectral and temporal modulations. Said sensitivity can be assessed using spectrotemporal modulation detection (STMD) tests. Although STMD tests are widely used, the effect of presentation level has been not been studied. We explored this effect by performing behavioral measurements on normal hearing (NH) participants and found that at higher ripple densities, STMD thresholds increased (i.e., worsened) with increasing level. The results of the model showed that the increased thresholds were caused by the detriment of the spectrotemporal representation at the AN level due to broadening of the cochlear filters and increased neural activity. Regression analysis revealed that the information at the AN was able to account for a large proportion of the behavioral data variance, supporting its value for STMD threshold predictions.

Finally, we developed frameworks to predict interaural time difference (ITD) discrimination thresholds. As a first step, we focused on NH and bilateral cochlear implant listeners. The model was validated by comparing its predictions with behavioral data from literature. It was able to qualitatively predict the effect of different stimulus parameters on ITD detection. The model predicted the data trends for unmodulated and low-frequency modulated stimuli, identifying the frequency regions of best performance, as well as the high frequency limit. In the mid-high modulation frequencies, the model underestimated the behavioral data.

On a second step, we focused on NH and hearing-impaired (HI) listeners. This model incorporated a decision device stage capable of yielding ITD threshold predictions in relevant (i.e., time) units. In a similar fashion, its predictions were validated with behavioral data from literature. In the NH case, the model was able to capture the behavioral data trends, although it had a tendency to overestimate the thresholds. In the HI case, the model was able to predict the behavioral trend at a group level. In both cases, additional model predictions revealed the underlying threshold functions. Additionally, we used the model to study the impact of hair cell loss on ITD discrimination in HI listeners. We found that for moderate levels of impairment, outer hair cell damage has a larger impact than inner hair cell damage at frequencies >500 Hz, partially explaining the variability of HI listeners.

The presented frameworks have a variety of practical applications, such as the development of new speech materials, reducing the time required for psychoacoustical tests, the assessment of speech processing algorithms, and the evaluation of stimuli parameters and/or signal processing schemes that improve ITD sensitivity.

Beknopte samenvatting

Om te begrijpen hoe geluid en spraak waargenomen worden, dient anatofysiologische kennis te worden aangevuld met psychoakoestische informatie. Het is echter een uitdagende taak om anatomische structuren en hun fysiologie te koppelen aan hun bijdrage aan auditieve verwerking.

Computationele modellen zijn waardevol om de kloof te dichten tussen de anatofysiologische kennis van het auditieve systeem en de psychoakoestiek van geluid en spraak. In de voorbije tientallen jaren, zijn fysiologisch-gebaseerde modellen steeds belangrijker geworden voor het onderzoeksdomein. Deze modelen simuleren de anatomische componenten en fysiologische processen die voorkomen in het auditieve systeem zo nauwkeurig mogelijk. Recentelijk, hebben Zilany et al. (2014, 2009) een model gepubliceerd van de auditieve periferie tot aan het niveau van de auditieve zenuw, dat gevalideerd is met een grote hoeveelheid fysiologische data uit de literatuur.

In deze thesis, hebben we dit model gebruikt als een *front end* in de ontwikkeling van verschillende *frameworks* om spraakverstaanbaarheid, spectrotemporale modulatie detectie (STMD) en perceptie van interaurale tijdsverschillend (ITD in het Engels) te bestuderen.

Eerst, hebben we met bekende objectieve maten spraakverstaan in ruis geëvalueerd en deze resultaten vergeleken met predicties van het *framework*. Spraak kan worden ontbonden in de envelope (relatief trage variaties in amplitude over tijd) en de temporele fijnstructuur (snelle oscillaties). Op woord niveau, vonden we dat fysiologisch geinspireerde maten die envelope-informatie gebruiken, het sterkst correleren met de gedragsmatige scores. Een multivariate lineare regressie analyse bevestigde dat deze maten spraakverstaan kunnen voorspellen. Op foneem niveau, vonden we dat foneemovergangen een consistente invloed hebben op de envelope component van spraak op het niveau van de auditieve zenuw.

Spraakverstaan is gebaseerd op de gevoeligheid van een individu voor spectrale en temporele modulaties. Deze gevoeligheid kan worden getest met spectrotemporele modulation detectie (STMD) taken. Hoewel deze taken veel worden gebruikt, is de invloed van presentatieniveau nog niet onderzocht. We hebben dit effect bestudeerd door bij normaalhorende deelnemers gedragsmatige STMD taken af te nemen en vonden dat voor hogere ripple densiteiten, STMD drempels stijgen (verslechteren) met stijgend presentatieniveau. De resultaten van het model toonden dat de stijging van de drempels veroorzaakt werd door een aangetaste spectrotemporele representatie op het niveau van de auditieve zenuw door het verbreden van de cochleaire filters en een toename in neurale activiteit. Regressie analyse bevestigde dat de informatie ter hoogte van de auditieve zenuw een groot deel van de variantie in de gedragsmatige metingen voorspelt, wat de relevantie van het model voor STMD drempels bevestigt.

Ten slotte, hebben we *frameworks* ontwikkeld die discriminatie drempels voor interaurale tijdsverschillen voorspellen. In een eerste deel hebben we ons toegelegd op normaalhorenden en personen die bilateraal geimplanteerd zijn met een cochleair implantaat. Het model werd gevalideerd door de predicties te vergelijken met gedragsmatige data uit de literatuur. Het model kan kwalitatieve voorspellingen maken over het effect van verschillende stimulus parameters op de detectie van ITD. Het model voorspelde de data trends voor ongemoduleerd en laag-frequent gemoduleerde stimuli. Zowel de frequentie gebieden met hoogste performantie als de hoogste frequentie limiet konden worden geïdentificeerd. Voor midden-hoge modulatiefrequenties, onderschatte het model de gedragsmatige data.

In een tweede stap, hebben we de focus verlegd naar normaalhorende en slechthorende luisteraars. Hier bevatte het model een stadium met een beslissingsapparaat dat ITD drempels kan voorspellen in een relevante (tijds-)eenheid. De predicties van dit model zijn op een gelijkaardige manier gevalideerd met gedragsmatige data uit de literatuur. In het geval van de normaalhorenden, kon het model de gedragsmatige trends voorspellen, al had het de neiging om de drempels te overschatten. Bij de slechthorenden, kon het model de trends op groepsniveau voorspellen. In beide gevallen, onthulden bijkomende voorspellingen van het model de onderliggende drempelfuncties. Daarnaast, gebruikten we het model om het effect van haarcelverlies op ITD discriminatie in slechthorenden te bestuderen. Voor matige niveaus van aantasting en voor frequenties boven 500 Hz, had schade aan de buitenste haarcellen meer invloed dan schade aan de binnenste haarcellen. Dit verklaart een deel van de variabiliteit in slechthorenden.

De voorgestelde *frameworks* hebben zeer verscheidene toepassingen, zoals de ontwikkeling van nieuwe spraakmaterialen, minimalisering van de duur van psychoakoestische testen, de evaluatie van spraakverwerkingsalgoritmes en onderzoek naar stimuli parameters en signaalverwerkingsalgoritmes die ITD perceptie verbeteren.

Contents

Acknowledgements iii

Abstract vii

Beknopte samenvatting xi

Contents xv

List of Figures xxi

List of Tables xxv

1 Introduction **1**

1.1 Motivation . . . 3

1.2 Computational Models . . . 4

1.3 The Peripheral Auditory System . . . 6

1.3.1 The Auditory Nerve . . . 7

1.3.2 Physiologically-inspired Models of the AN . . . 12

1.4 Objective & Outline . . . 16

2 Predicting Phoneme and Word Recognition in Noise **19**

2.1 Abstract . . . 21

2.2 Introduction . . . 22

2.3 Background . . . 25

2.3.1 Neurogram-based Metrics . . . 25

2.3.2 Filterbank-based Metrics . . . 27

2.4 Materials & Methods . . . 30

2.4.1 Speech Material . . . 30

2.4.2 Objective Measures . . . 32

2.4.3 Behavioral Measurements . . . 35

2.4.4 Comparison . . . 36

2.5 Results . . . 37

2.6 Discussion . . . 41

2.7 Conclusions . . . 44

3 The Effect of Level on Spectrotemporal Modulation Detection 45

3.1 Abstract 47

3.2 Introduction 48

3.3 Behavioral Measurements 49

3.3.1 Materials & Methods 49

3.3.2 Results 53

3.4 Computational Model 55

3.4.1 Stimuli 56

3.4.2 AN Model 56

3.4.3 Neurogram activity 58

3.4.4 Neurogram frequency profiling 58

3.4.5 Dispersion 59

3.4.6 Regression 60

3.5 Discussion 63

3.6 Conclusions 65

4 ITD Discrimination in Acoustical and Electrical Hearing 67

4.1 Abstract 69

4.2 Introduction 70

4.3 Materials & Methods 73

4.3.1 Stimuli 74

4.3.2 Peripheral Model 78

4.3.3 Shuffled Cross Correlogram 79

4.3.4 Binary Classifier Characterisation 80

4.3.5 Model Performance Evaluation 83

4.4 Results 84
4.4.1 Acoustical Hearing 84
4.4.2 Electrical Hearing 90
4.5 Discussion 93
4.5.1 Model Purpose and Assumptions 93
4.5.2 Performance Assessment and Future Steps 94
4.6 Conclusions 97

5 ITD Discrimination in Normal and Hearing Impaired Listeners 99

5.1 Abstract 101
5.2 Introduction 102
5.3 Framework Description 104
5.3.1 Auditory Periphery 105
5.3.2 Coincidence Detection 107
5.3.3 Decision Device 111
5.4 Model Results 113
5.4.1 Framework Validation 113
5.4.2 Contribution of OHC/IHC Loss to HI 119
5.5 Discussion 122
5.5.1 Framework Structure 122
5.5.2 Framework Validation 124
5.5.3 Contribution of OHC/IHC Loss to HI 126
5.5.4 Comparison with Existing Models 127
5.5.5 Framework Limitations 129
5.6 Conclusions 131

6 General Discussion and Conclusion 133
6.1 Core Findings . . . 135
6.2 Applications . . . 137
6.3 Future Outlook . . . 139
6.4 General Conclusion . . . 141
References 143
Scientific Acknowledgements 165
Personal Contribution 167
Conflict of Interest Statement 169
Curriculum Vitae 171
List of Publications 173

List of Figures

1.1 Basic structure of the human peripheral auditory system 8

1.2 Frequency tuning curves from guinea pig AN fibers 9

1.3 Block diagram of the Zilany et al. (2014, 2009) model 15

2.1 Block diagram 30

2.2 Example segmentation of *bot* 31

2.3 Example ENV and TFS neurograms of *bot* 33

2.4 Example ES neurograms and cortical representation of *bot* . . . 34

2.5 Objective measures of the complete set averaged across words . 39

2.6 Phoneme scores vs objective measures averaged across participants at a word level 40

2.7 Correlation between the behavioral scores and the computed metrics (per phoneme) 40

3.1 Spectrograms of the spectrotemporally modulated stimuli 51
3.2 STMD thresholds 54
3.3 STMD thresholds with a ripple density of 4 ripples/oct 54
3.4 Block diagram of the proposed model 55
3.5 ESNs of spectrotemporally modulated stimuli 57
3.6 Neurogram activity 58
3.7 Frequency profiles of the ESNs 59
3.8 ESNs dispersion 61
3.9 ESN representation at (behavioral) threshold 61
3.10 Exponential regression model 62
3.11 Model predictions of the behavioral thresholds across levels . . 62

4.1 Block diagram of the proposed model 74
4.2 Shuffled Cross Correlogram (SCC) computation. 80
4.3 SCC curves area computation and mismatch 81
4.4 Binary Classifier Characterization (BCC) computation. 83
4.5 Behavioral data and model predictions for pure tones 84
4.6 Behavioral data and model predictions for SAM tones 86
4.7 Behavioral data and model predictions for transposed tones . . 86
4.8 Comparison of behavioral data and model predictions for pure, SAM, and transposed tones 87
4.9 Behavioral data and model predictions for raised sine with variable modulation frequency 88
4.10 Behavioral data and model predictions for raised sine with variable modulation depth 89
4.11 Behavioral data and model predictions for trapezoid modulation 89

4.12 Behavioral data and model predictions for unmodulated electric pulses . . . 91
4.13 Behavioral data and model predictions for electric SAM pulses 92
4.14 Behavioral data and model predictions for electric trapezoid modulation pulses . . . 93

5.1 Block diagram of the proposed model . . . 104
5.2 Shuffled Cross Correlogram (SCC) computation . . . 108
5.3 Different stages of the Coincidence Detection block for a 1 kHz pure tone . . . 110
5.4 Computation distributions for a 1 kHz pure tone . . . 111
5.5 Neurometric function for a 1 kHz pure tone . . . 112
5.6 Validation of the predicted ITD thresholds for pure tones in NH listeners . . . 116
5.7 Validation of the predicted ITD thresholds for bandpass noise in NH listeners . . . 117
5.8 C_{OHC} and C_{IHC} values for auditory thresholds for two CFs . . 119
5.9 Validation of the predicted ITD thresholds for bandpass noise in HI listeners . . . 120
5.10 Dispersion plots of predicted vs behavioral ITD thresholds for bandpass noise stimuli . . . 121
5.11 Effect of attributing different auditory threshold shift proportions to OHC/IHC impairment for bandpass noise . . . 121

List of Tables

2.1 Multivariate linear regression models results 38

4.1 NH stimuli parameters 75
4.2 BiCI stimuli parameters 77

5.1 Participants characteristics and stimuli parameters 114

CHAPTER 1

Introduction

1.1 Motivation

Our current knowledge of the human auditory system has been built using a variety of different sources. Anatophysiological studies (either *in vivo* or *ex vivo*) have taught us which elements compose it, how they work, and how sound is transmitted and processed from the outer ear all the way to the brain.

Electrophysiological recordings (e.g., electroencephalography, EEG) and imaging techniques (e.g., functional magnetic resonance imaging, fMRI) have complemented the anatophysiological studies. They have provided insight into *how* and *where* in the brain are sounds processed.

We have obtained most of our knowledge on how we perceive sounds (including speech) from behavioral experiments. In these, a stimulus with specific characteristics is presented to a participant. Then, the participant is asked to report what (s)he heard or to discriminate between a reference and a target condition.

These approaches have been fundamental in the study and understanding of the human ear and hearing. However, they present a couple of drawbacks. Due to the invasiveness of the anatophysiological techniques, these have been performed mostly either on humans *post mortem* (e.g., Spoendlin and Schrott, 1988) or on experimental mammals such as chinchillas, gerbils, guinea pigs, and cats (e.g., Liberman, 1978; Müller and Robertson, 1991; Young and Sachs, 1973). Non-invasive techniques present some inherent disadvantages. For instance, EEG is only able to provide the average response of a large neural population, while fMRI has a poor temporal resolution (in the order of seconds, Kim et al., 1997). The evaluation of behavioral measurements is a very resource-consuming task. Participants recruitment can be specially difficult if they need to have specific characteristics. Depending on the number of phenomena or conditions to be studied, measurement sessions can be very long, which might affect participants' concentration and, therefore, their performance. Additionally, participants are often required to attend numerous sessions. Furthermore, behavioral measurements cannot provide a direct insight into the responses of the involved elements along the auditory pathway.

The technological advances from the last decades have reduced the cost and exponentially increased the performance of computers. This has allowed computational models to emerge as valuable tools for the audiology community that function as a complement to the aforementioned approaches.

In this thesis, we developed different computational models to investigate a variety of auditory processes in different types of listeners. We focused on speech intelligibility, spectrotemporal modulation detection, and interaural time difference sensitivity. These models share and take advantage of a physiologically-inspired front end in the form of the model of the auditory periphery proposed by Zilany et al. (2014, 2009).

1.2 Computational Models

A complete, thorough comprehension of the human auditory system and sound perception requires both an anatophysiological and a behavioral/psychoacoustical approach, since neither physiological nor perceptual data can provide sufficient information on their own. On one hand, physiological studies are unable to identify the function of the studied structure. On the other, perceptual studies are unable to identify the implementation of these functions (Delgutte, 1996). However, it can be hard to reconcile these two domains. Relating an anatomical structure and its physiology to its contribution to auditory perception is, more often than not, a challenging task.

Computational models are valuable tools that aim to bridge the gap between the physiology of the auditory system and the psychoacoustics of sound and speech. In other words, they can help us explain different auditory phenomena psychoacoustically based on our knowledge of the auditory system anatophysiology. They are capable of integrating and synthesizing the discoveries of anatomists, physiologists, health care professionals, psychophysicists, and engineers, among others, in a single, coherent framework.

Additionally, computational models present several advantages that make them very attractive for studying human auditory phenomena (Meddis et al., 2010). For instance, they can provide a detailed description of the structure(s) of interest. The level of detail will depend on its specific purpose. They are quantitative, since their (numeric) parameters must be defined before they can be run. They can be versatile and flexible, allowing to explore the effects of a wide range of values and conditions of the desired parameters, limited only by the availability of computational time and power. Usually, computational models are defined as code of a computational program, making them portable: they can be shared easily and run on different computers (provided that these computers have the minimum technical requirements). Lastly, they are usually modular, potentially making it possible to integrate models of specific processes into larger, more comprehensive models. In this way, they can help determining how the deficit of one (or more) components affects the overall system performance.

One of the advantages of computational models is their versatility. Not only can we choose *what* we want to model (e.g., values, parameters, conditions), but we can also decide *how* we want to model it.

There is a long story of different modelling paradigms of different parts of the auditory system (e.g., transmission line models, Duifhuis (2004); Verhulst et al. (2012); electrical models, Mountain and Hubbard (1996); mechanical models, De Boer (1996); Hubbard and Mountain (1996); electromechanical models, Rosowski (1996); Steele et al. (1993); finite element models, Cheng (2007)). However, the trend of the last couple of decades is to simulate as closely as possible the anatomical components and the physiological processes that occur in the auditory system (Meddis et al., 2010). Although this type of models is usually computationally expensive, it presents valuable advantages. For example, it provides a more transparent approach to comprehend the auditory system and its different processes. Including physiological information is essential when modelling the effects of impairment or deficiencies of specific structures. Model parameters come (mostly) from physiology itself, therefore reducing the risk of overfitting. Furthermore, incorporating anatophysiological mechanisms in a model allows a straightforward comparison of its output with its biological counterpart. In the end, only models that simulate the auditory anatophysiological aspect closely are the ones that can be verified directly using actual physiological measurements.

1.3 The Peripheral Auditory System

Figure 1.1 shows an schematic of the human peripheral auditory system. Basically, it consists of three sections: the outer, middle, and inner ear. Each of them is briefly explained below[1], together with a short description of how they are commonly modelled[2].

The first element of the outer ear is the pinna. It helps capturing the sound waves, which are reflected and attenuated before entering the ear canal (also known as meatus). The sound travels down the ear canal until it reaches the eardrum, which marks the beginning of the middle ear. The effect of the pinna, the ear canal, the head, and sometimes of the body are usually modelled using a head-related transfer function (HRTF). Some frameworks reproduce the HRTF features by modelling the interaction of the sound with the aforementioned body parts (e.g., Lopez-Poveda and Meddis, 1996; Walsh et al., 2004), while some others model the HRTF using a cascade of digital filters (e.g., Kistler and Wightman, 1992; Kulkarni and Colburn, 2004).

The sound waves cause the eardrum (also known as tympanic membrane) to vibrate. These vibrations are transmitted through the middle ear by three bones known as the ossicles: malleus, incus, and stapes (also known as hammer, anvil, and stirrup, respectively). They are responsible for converting the sound wave into mechanical motion and to transmit the latter to the oval window. The middle ear has been modelled using different approaches. For instance, it has been modelled using (biomechanical) finite element schemes (Sun et al., 2002). Some other frameworks have modelled it as electrical circuits thanks to the analogy between electrical and acoustical systems (Rosowski, 1996). However, the most popular method consists in modelling it using one (or more) linear digital filters with an appropriate frequency response (Holmes et al., 2004; Tan and Carney, 2003).

The oval window divides the middle ear from the inner ear. The inner ear consists of the semicircular canals and the cochlea. The former are part of the vestibular system and are responsible for spatial orientation and equilibrium. However, they play no role in hearing. Therefore, we will focus on the cochlea.

---

[1] For a more extensive and detailed explanation, see Moore (2013)

[2] For a more extensive and detailed description, see Meddis et al. (2010)

The cochlea is a spiral-shaped organ filled with almost incompressible lymphatic fluids. The start of the cochlea (where the oval window is situated) is known as the base, while the end of the cochlea (the tip) is known as the apex. It is divided along its length by two membranes: the tectorial membrane and the basilar membrane (BM). Between these two membranes are hair cells with stereocilia on their top, all of which form part of a structure known as the organ of Corti. Those hair cells on the side of the arch close to the outside of the cochlea are known as outer hair cells (OHCs), while the opposite ones are known as inner hair cells (IHCs). The latter are responsible for converting the mechanical movements into neural activity as follows. When the stapes moves the oval window, a pressure difference is applied to the BM, causing a shearing motion between the BM and the tectorial membrane. This motion shifts the IHC stereocilia, which opens transduction channels and allows ions into the hair cell. Then, the IHC is depolarized and releases neurotransmitter into the hair-cell-neuron synapse. The neurotransmitter travels to the receptors on the auditory nerve (AN), generating an action potential. The IHC activity in response to BM excitation is commonly modelled phenomenologically with a non-linear gain followed by a low-pass filter (e.g., Robert and Eriksson, 1999; Zhang et al., 2001). Alternatively, it has also been modelled using electrical circuit analogs of the whole organ of Corti (e.g., Lopez-Poveda and Eustaquio-Martín, 2006).

1.3.1 The Auditory Nerve

The auditory nerve (AN) is an interesting and reasonable starting point for modelling efforts of auditory phenomena, since all the information elicited by acoustic stimuli that reaches the brain is encoded here (Delgutte, 1996). Furthermore, the AN has received a great deal of attention from the scientific community, resulting in an extensive amount of information regarding the properties of the individual components.

Due to its relevance to this thesis, in this section we provide a summary[3] of the AN's anatomy, physiology, and of its response properties to acoustic stimulation. Understanding its basic structure and functional organization is the first step towards comprehending its computational counterpart. Then, in Sec. 1.3.2 we present an overall view of the most relevant physiologically-inspired models of the AN.

---

[3] For a more extensive and detailed description, see Møller (2000); Ruggero (1992).

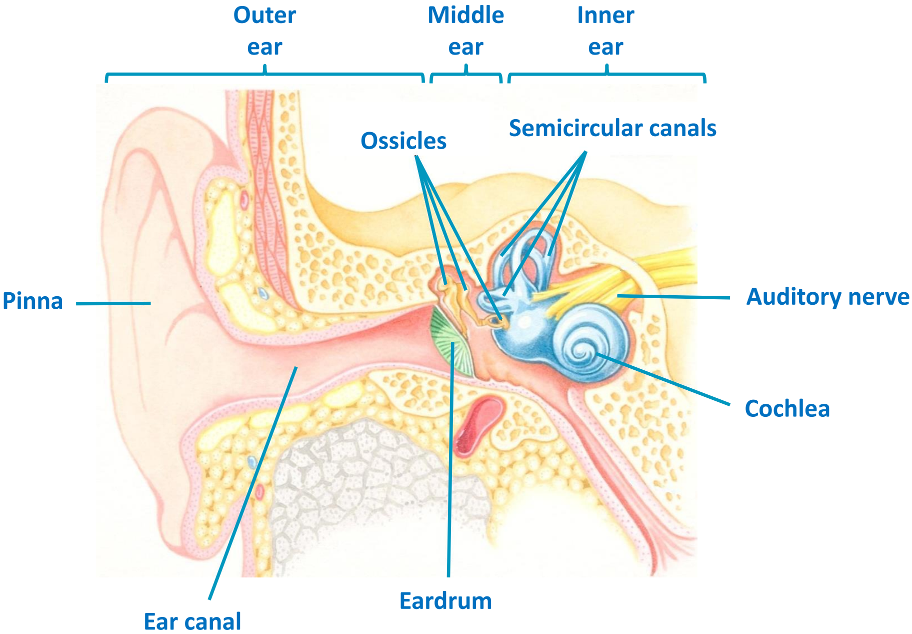

Figure 1.1: Basic structure of the human peripheral auditory system. Adapted from Sergent (2018).

Anatomy

The human AN consists of approximately 30,000 fibers (Spoendlin and Schrott, 1989), which can be classified into two types. On one hand, type I fibers constitute about 95% of the AN (Liberman, 1982*b*). These fibers innervate IHCs and have been identified as afferent, which means that they send information from the auditory periphery to the central auditory system. On the other hand, type II fibers constitute the remaining 5%. These fibers innervate OHCs and has been speculated that they are efferent, meaning that they modulate the functioning of the peripheral system. However, this is still debated and is a topic of ongoing research (Weisz et al., 2009; Zhang and Coate, 2017). Therefore, we will focus only on type I fibers.

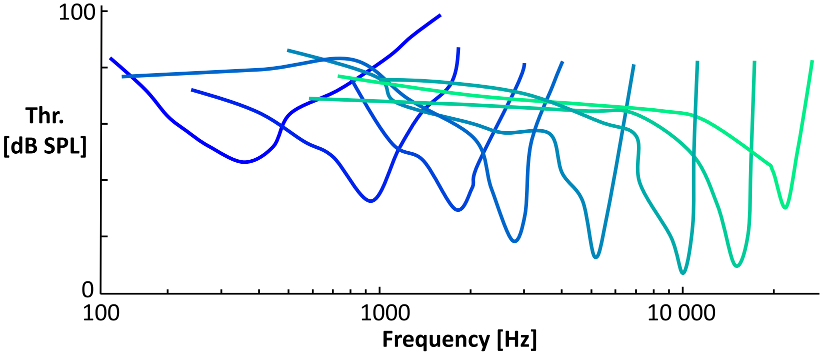

Figure 1.2: Frequency tuning curves from guinea pig AN fibers. Adapted from Evans (1972).

Physiology

An important property of AN fibers is their spontaneous rate (SR). We can define a fiber's SR as the discharge rate (in spikes/s) measured when no external stimulus is present. Based on this, AN fibers can be classified in three different types[4](Liberman, 1978; Winter et al., 1990). High spontaneous rate (HSR) fibers discharge at a rate of >15 spikes/s and comprise ∼60% of the population. They have dynamic ranges between 30 and 40 dB. Medium spontaneous rate (MSR) fibers discharge at a rate between 0.5 and 15 spikes/s and comprise ∼20% of the population, while low spontaneous rate (LSR) fibers discharge at a rate of <0.5 spikes/s and comprise the remaining ∼20% of the population. MSR and LSR fibers have higher dynamic ranges (between 50 and 60 dB).

Another critical characteristic of all AN fibers is their capability of responding selectively to pure tones with different frequencies. The sound pressure level (SPL) needed to increase the fibers' rate responses above their SRs (typically 20%) depends on the stimulus frequency. This relation can be represented graphically in the so-called frequency tuning curves. Figure 1.2 shows example frequency tuning curves obtained from guinea pig AN fibers (Evans, 1972). The frequency at which a fiber is the most sensitive (corresponding to the tip of its respective tuning curve) is known as the fiber's characteristic frequency (CF, Kiang, 1965).

---

[4] Some authors (e.g., Evans and Palmer, 1980) consider only two types: HSR and LSR.

Usually, the range of CFs extends roughly over the frequency range of hearing. There is an explicit mapping between the CF of a fiber and its point of cochlear innervation (Liberman, 1982*a*). Fibers with low CFs are situated closer to the cochlea's apex, while fibers with high CFs are located closer to the cochlea's base. Interestingly, this distribution can be described mathematically in a very similar way in most mammals, differing only in the frequency range (Greenwood, 1961, 1990).

Therefore, so far we can understand the AN in two important dimensions. The first one corresponds to its level of (spontaneous) activity, while the second one corresponds to frequency (in other words, its cochlear location). Additionally, the AN has certain physiological response properties whose study has been fundamental in the development of computational models (Lopez-Poveda, 2005). These phenomena are briefly described as follows.

Adaptation This term comprises different properties (Westerman, 1985). First, AN adaptation refers to response recovery from previous stimulation. Instantly after the stimulus ends, the discharge rate of the AN fiber falls to its lowest (well below its spontaneous rate). Afterwards, it increases gradually with time until it reaches its spontaneous rate again. AN adaptation also refers to the property in which a sudden increment/decrement of the stimulus level causes an increment/decrement in discharge rate. Lastly, AN adaptation refers to the phenomena in which the fiber's discharge rate has a sudden increase at the stimulus's onset and then decays over time until it becomes stable (i.e., until it is adapted). This decay is faster during the first 10 to 20 ms of the stimulus.

Phase-Locking When there is no external stimulus, AN fibers fire at their SR. However, in presence of a sinusoidal stimulus, the spike timing depends on the stimulus frequency (Rose et al., 1967). In the case of low frequencies, AN fibers tend to fire during the positive half-cycle of the sinusoid. Although phase-locking can be seen in AN fibers regardless of their CF, the degree of phase-locking decreases with increasing stimulus frequency (Johnson, 1980). It has been shown that this reflects the low-pass filtering properties of the IHCs (Palmer and Russell, 1986).

Suppression This phenomenon occurs when the response of an AN fiber to a given sound (called the excitor) decreases in the presence of a second sound (called the suppressor) at an appropriate level and frequency. Interestingly, suppression may be caused even by suppressors that do not excite the fiber themselves. Usually, two-tone suppression is expressed as the increment of level (in dB) of the excitor tone required to maintain a certain discharge rate while the suppressor is being played (Javel et al., 1978).

Level-Dependent Frequency Tuning & Best Frequency Shifts If we plot the discharge rate of AN fibers as a function of stimulus frequency, we would obtain a curve with a peak at the fiber's CF. As we increase the level of the stimulus, this peak broadens, which leads to reduced frequency selectivity (Müller and Robertson, 1991; Rose et al., 1971). This is due to the saturation of the fiber's discharge rate at high levels and, most importantly, the progressive widening of the BM response with increasing level.

The best frequency (BF) is the frequency to which a fiber responds the most at any given level. When the stimulus level is near a fiber's threshold, the CF is the BF. However, in this case, increasing the level shifts the BF away from the fiber's CF. The direction of this shift depends on the fiber's CF. It has been suggested that this shift reflects the BM response properties (Carney, 1999; Rose et al., 1971).

Level-Dependent Phase Within the phase-locking range, the AN activity tends to synchronize with the stimulus frequency. However, the time at which the spikes are generated differs from the time of the stimulus waveform peak. This phase shift increases with increasing frequency. Furthermore, the phase shift also depends on the level of the stimulus. For frequencies below the CF, the fiber discharges increasingly later in the cycle with increasing level. For frequencies above the CF, the opposite happens (Anderson et al., 1971). Similarly to the BF shifts, it has been suggested that the level-dependent phase shifts are associated with the level-dependent bandwidth of the BM and its phase changes (Geisler and Rhode, 1982).

1.3.2 Physiologically-inspired Models of the Auditory Nerve

In this section we provide a brief, non-exhaustive overview[5] of model families that are important because of their historic relevance or because of their contribution to the state-of-the-art models.

One of the first attempts to predict auditory perception based on the AN activity was done by Siebert (1965, 1968, 1970). This model incorporates key features of the activity patterns of the AN, such as frequency tuning, saturating rate-level functions, and a logarithmic cochlear map (Kiang, 1965). He combined this AN information with signal detection theory (Green and Swets, 1966) in the form of Poisson statistics. In this model, an increase in the level causes a constant increase in the number of recruited AN fibers. However, at the same time, an equal number of AN fibers becomes saturated. Therefore, the number of fibers that effectively convey level information remains constant. He deduced that the level was coded by the total number of active fibers (rather than the amount of activity) at a specific place in the AN and concluded that the fact that excitation patterns of the AN are not identical for repetitions of the same stimulus limited a listener's performance when detecting differences between acoustic stimuli. Based on this principle, the model was able to predict pure tone intensity discrimination according to Weber's law[6]. The importance of this model relies on its pioneer approach in using a systematic methodology for predicting psychoacoustical (i.e., behavioral) performance based on AN activity patterns.

Deng and Geisler (1987) proposed a model comprising four stages. The first one is a middle-ear filter. Then, the output of this stage is fed to a stage that reproduces the response of the cochlea using the difference equations proposed by Viergever (1980). Their parameters are tuned to account for AN threshold curves with different CFs. This stage has two versions, a linear and a non-linear one. The next stage mimics the IHC function using a sigmoid-type non-linearity and a low-pass filter afterwards, converting velocity into IHC receptor potential. The last stage accounts for the synaptic effects and converts the IHC potential into AN fiber firing probability. This model is able to reproduce level-dependent

---

[5] For a more extensive review, see Delgutte (1996); Lopez-Poveda (2005); Meddis et al. (2010)

[6] Weber's law states that the just noticeable change in a stimulus is a constant proportion of the original one. It has successfully characterized human responses to a wide variety of sensory stimuli (Deco et al., 2007).

tuning of HSR fibers. Although it is technically capable of modelling adaptation and suppression, its predictions for these phenomena have not been validated with physiological data. The authors concluded that the non-linear version of the model was able to better reproduce the responses of the AN fibers to speech. However, due to the automatic gain control of the synaptic stage, the IHC non-linearity had little influence when producing realistic outputs.

An important predecessor of many models is the gammatone filter (Aertsen et al., 1980; De Boer, 1975; De Boer and De Jongh, 1978). It was developed to mimic the impulse response of AN fibers using reverse correlation techniques. The impulse response of the gammatone filter consists of the product of a carrier tone with a frequency equal to the fiber's BF and a gamma distribution that determines the shape of the response envelope. Given its relatively simple and efficient computational implementation (Slaney et al., 1993), it has been widely used to model auditory frequency selectivity, as well as to simulate the cochlear excitation pattern. However, the gammatone filter has a symmetric frequency response, making it inadequate to simulate the asymmetric response of the BM.

There have been attempts to make more physiologically accurate versions of the gammatone filter. Inspired by the fact that the BM and AN fibers responses are frequency modulated, Irino and Patterson (1997, 2001) proposed the gammachirp filter. This model's version consisted of three stages: a gammatone filter, a low-pass, and a high-pass filter (the latter with a level-dependent cutoff frequency). Because of its asymmetric frequency response, this model's impulse response shows a chirp, hence its name. Although it has been used to design filterbanks that reproduce the response of human auditory filters over a large range of levels and frequencies, it is incapable of reproducing physiological data of BF shifts with level and other nonlinear phenomena, such as two-tone suppression, level-dependent phase responses, and combination tones.

The model of Robert and Eriksson (1999) simulates the response of the auditory periphery in response to arbitrary sounds. It is composed of four stages. First, linear bandpass filters simulate the frequency response of the outer and middle ear. The next stage simulates the cochlea by using a filter bank with each filter representing a specific location of the BM. This stage is of particular interest. Each element of the filterbank consists of two filters: a passive one and an active one, with time-independent and time-dependent parameters, respectively. Both are implemented as gammatone filters (Lyon, 1997). A feedback process controls the active filter's tuning and gain (Carney, 1993). Interestingly, this feedback control of each filter depends not only on its output, but also on the outputs of neighboring filters. Then, the output of each cochlear filter is used as an input to the computational model of the IHC-AN synapse proposed by Meddis (1986, 1988). This stage provides the probability of the AN fiber firing. The last stage converts this probability into actual spikes, taking into account absolute and relative refractory periods. This model is capable of reproducing adaptation, rate-intensity functions of the LSR and HSR fibers, and responses to pure tones in noise. However, it cannot account for effect of level and frequency on phase-locking, as well as level dependent phase.

Later, Sumner et al. (2002, 2003) proposed a model based on the work of Lopez-Poveda and Meddis (1996); Lopez-Poveda et al. (1998); Meddis (1986, 1988) comprising five stages. First, the input stimulus is passed through an outer- and middle-ear bandpass filter, yielding as an output a signal equivalent to the stapes velocity. The latter is then fed to a so called dual-resonance non-linear filter, which consists of two parallel pathways (one linear and one non-linear). Their outputs are summed, yielding as an output the BM velocity. The next stage, based on the biophysical model proposed by Shamma et al. (1986), simulates the transduction at the IHC level. Afterwards, the model accounts for the calcium kinetics in the release of neurotransmitter in the synaptic connection. The last stage generates the action potentials of the AN based on the quantal (i.e., discrete) and stochastic release of neurotransmitter. This model is capable of simulating rate-level functions for fibers with all three different SRs. It is also able to reproduce response properties of the AN such as level-adaptation and phase-locking, but fails to account for suppression growth rates as well as long-term response properties (e.g., offset response).

The Zilany et al. (2014, 2009) Model

Zilany et al. (2009) proposed a model of the physiological response at the AN level. It builds upon the models proposed by Bruce et al. (2003); Carney (1993); Zhang et al. (2001); Zilany and Bruce (2006, 2007). They implemented an IHC-AN synapse adaptation that includes both exponential and power-law dynamics.

This model is comprised of various modules, each mimicking a particular function of the auditory periphery. Its block diagram is shown in Fig. 1.3 and briefly explained as follows.

First, the stimulus is passed through a filter emulating the middle ear. The output is then passed through a signal path and a control path. The former simulates the behavior of the OHC-controlled filtering properties of the BM in the cochlea and the transduction properties of the IHCs by a succession of non-linear and low-pass filters. The latter simulates the function of the OHCs in controlling BM filtering. The output of the control path feeds back into itself and into the signal path, as well. The IHCs output then goes through an IHC-AN synapse module: an exponential process that drives two parallel power-law adaptation paths in order to account for slow and fast adaptations. Instead of tuning the parameters of these two power-law functions *ad hoc*, they were kept fixed and were not optimized to fit individual AN responses (avoiding overfitting).

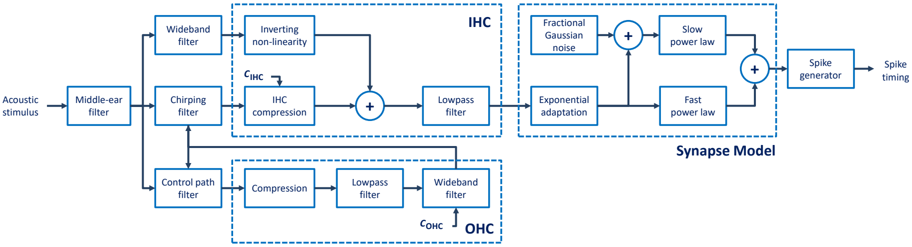

Figure 1.3: Block diagram of the Zilany et al. (2014, 2009) model. The input to the model is the pressure waveform of an acoustic stimulus. Its output is a series of AN spike times. C_{OHC} and C_{IHC} are scaling constants that reflect the effect of OHC and IHC impairment, respectively (Sec. 5.4.1). Adapted from Zilany et al. (2014, 2009).

One of the most attractive properties of this model is its capability of modelling the effects of OHC and IHC impairment using two scaling constants: C_{OHC} and C_{IHC} (Bruce et al., 2003; Zilany and Bruce, 2006), respectively. C_{OHC} is introduced at the output of the control path, while C_{IHC} is introduced in the signal path. In both cases, a value of 1 represents normal function, while values closer to 0 represent larger impairment.

In its next iteration, Zilany et al. (2014) readjusted the parameters of the synapse model to improve the simulation of physiological discharge rates at saturation for higher CFs, as well as to correct their response to low frequency tones.

The responses of the Zilany et al. (2014, 2009) model have been validated with physiological data from literature over a wider range than previously existing ones. These include spontaneous activity (from three different AN fiber types, Sec. 1.3.1), responses to pure tones (including pure tones under the increment/decrement paradigm), forward masking, responses to amplitude-modulated tones, and responses to noise.

This has allowed it to be successfully used to study a variety of auditory phenomena, such as neural adaptation to sound level (Zilany and Carney, 2010), sensory responses to musical consonance-dissonance (Bidelman and Heinz, 2011), overshoot adaptation (Jennings et al., 2011), masking release (Bruce et al., 2013), frequency selectivity (Jennings and Strickland, 2012), and neural coding of chimaeric speech (Heinz and Swaminathan, 2009), among others. However, many other auditory processes could be modelled using this framework and benefit greatly from its physiologically-inspired nature.

1.4 Objective & Outline

The aim of this thesis was to study a number of auditory phenomena using computational models which benefit from implementing a physiologically-based approach. Specifically, we used the physiologically-inspired model of the auditory periphery at the AN level proposed by Zilany et al. (2014, 2009) (Sec. 1.3.2) as a core of various computational frameworks to study different auditory processes, namely:

Speech Intelligibility

SI is one of the most important outcome measures of audiological rehabilitation. Several objective measures have been used to predict SI. However, most of them incorporate little biological aspects about the auditory periphery. In Ch. 2, we assessed SI using different filterbank-based metrics and compared their performance with the physiologically-inspired framework we developed.

Modulation Detection

Speech understanding relies (at least partially) on an individual's sensitivity to spectral and temporal modulations. Said sensitivity can be assessed using spectrotemporal modulation detection (STMD) tests. In these tests, participants are asked to discriminate between a modulated and non-modulated stimulus. The threshold is defined as the minimal modulation depth that the participant can detect. Although STMD tests are widely used, the effect of presentation level has been understudied. In Ch. 3, we explored this effect by performing behavioral measurements on NH participants and we used a physiologically-inspired model to expand on them.

Binaural Hearing

Sound localization depends on the perception of binaural cues: interaural level differences (produced by the sound arriving softer to the ear that is farthest from the sound) and interaural time differences (ITDs, produced by the sound arriving later to the ear that is farthest from the sound). In this thesis, we focused on the latter. In Ch. 4, we developed a model to predict ITD discrimination thresholds for NH and electrical stimulation (through bilateral cochlear implants). In Ch. 5, we developed a physiologically-based model to investigate ITD thresholds of normal and hearing impaired listeners. In both cases, we validated the models by comparing their predictions against behavioral data reported in literature.

All the developed models presented in the different chapters are published or in press for publication in international peer-reviewed scientific journals.

CHAPTER 2

Predicting Phoneme and Word Recognition in Noise[1]

---

[1] The work presented in this section has been published as Moncada-Torres, A., van Wieringen, A., Bruce, I. C., Wouters, J. and Francart, T. (**2017**). “Predicting phoneme and word recognition in noise using a computational model of the auditory periphery,” The Journal of the Acoustical Society of America **141**(1), 300–312. Changes are limited to layout, graphical appearance, and minor editing.

2.1 Abstract

Several filterbank-based metrics have been proposed to predict speech intelligibility (SI). However, they incorporate little knowledge of the auditory periphery. Neurogram-based metrics provide an alternative, incorporating knowledge of the physiology of hearing by using a mathematical model of the auditory nerve response. In this work, SI was assessed utilizing different filterbank-based metrics (the Speech Intelligibility Index and the Speech-based Envelope Power Spectrum Model) and neurogram-based metrics, using the biologically-inspired model of the auditory nerve proposed by Zilany et al., [2009, *The Journal of the Acoustical Society of America*, **126**(5), 2390–2412] as a front-end and the Neurogram Similarity Metric and Spectro Temporal Modulation Index as a back-end. Then, the correlations with behavioral scores were computed. Results showed that neurogram-based metrics representing the speech envelope showed higher correlations with the behavioral scores at a word level. At a per-phoneme level, it was found that phoneme transitions contribute to higher correlations between objective measures that use speech envelope information at the auditory periphery level and behavioral data. The presented framework could function as a useful tool for the validation and tuning of speech materials, as well as a benchmark for the development of speech processing algorithms.

2.2 Introduction

Speech intelligibility (SI) is assessed using behavioral or objective measures. Usually, in the former, a group of participants are asked to listen to a stimulus under particular conditions and asked to register or identify what they heard (Miller, 2013). Such tests are used to characterise a patient's hearing and to evaluate the performance of new hearing devices. Furthermore, behavioral measures have allowed gathering valuable information on auditory perception. Objective measures are a complementary approach. They are also used to evaluate instruments' performance, since they present several advantages on their own. Their parameters can be set and tuned flexibly to investigate different conditions. Additionally, they can be obtained faster and in an automated way using a computer. Several objective measures have been used to predict SI. For the purposes of this paper, we will divide them into two groups: *filterbank-based* metrics and *neurogram-based* metrics.

On one hand, filterbank-based metrics model the frequency selectivity of the auditory periphery by separating the speech signal into various frequency bands. There are several well established objective measures that have this working principle at their core, such as the Articulation Index (AI, French and Steinberg, 1947), the Speech Transmission Index (STI, Steeneken and Houtgast, 1980), the Speech Intelligibility Index (SII, ANSI, 1997, Sec. 2.4.2), and the Speech-based Envelope Power Spectrum Model, (sEPSM, Jørgensen and Dau, 2011, Sec. 2.4.2). These metrics have been moderately successful in predicting SI of normal-hearing (NH) listeners under various conditions (e.g., Bradley, 1986; Jørgensen et al., 2013; Kryter, 1962; Pavlovic, 1987). However, their approach for modelling the auditory periphery can be thought of as simplistic, since they base their SI prediction on acoustic features within each band rather than on knowledge of any aspect of physiological processing performed by it. Furthermore, taking into account the anatomical and the physiological mechanisms underlying the auditory system can provide a better understanding of the different factors that affect SI of NH or hearing-impaired (HI) listeners. For example, incorporating biological information is essential when modelling effects of hearing impairment. A physiologically-based approach allows to incorporate different impairment conditions at various stages, e.g., sensorineural hearing loss due to damage to the inner hair cells (IHCs) or outer hair cells (OHCs).

On the other hand, neurogram-based metrics use mathematical models to mimic the physiological response of the auditory periphery. They represent neural activity as a function of characteristic frequency (CF) and time (i.e., the neurogram itself). These metrics try to predict SI with little influence of higher order processes (e.g., cognitive, linguistic, phonetic; Sidwell and Summerfield, 1986) by comparing neurograms of clean and corrupted speech. Since they take into account anatomy and physiological processes, we believe they can provide a better understanding of underlying factors in the auditory system that affect SI.

Different physiological models have been developed in the past to investigate different phenomena, such as pitch and timbre (Lyon and Shamma, 1996), the responses of high-spontaneous-rate auditory nerve (AN) fibers (Zhang et al., 2001), and to predict neural activity to speech (Bruce et al., 2003), for instance. Recently, Zilany et al. (2009) proposed a model of the physiological response at the AN level (Sec. 2.3.1). Its responses have been validated with physiological data over a wider dynamic range than previously existing models. This model has been successfully used to study a variety of auditory phenomena, such as neural adaptation to sound level (Zilany and Carney, 2010), sensory responses to musical consonance-dissonance (Bidelman and Heinz, 2011), overshoot adaptation (Jennings et al., 2011), masking release (Bruce et al., 2013), frequency selectivity (Jennings and Strickland, 2012), and neural coding of chimaeric speech (Heinz and Swaminathan, 2009). However, its use for assessing SI and the benefits of its biologically-inspired nature have been evaluated only in a limited number of studies so far. For example, Hines and Harte (2012) used it together with their Neurogram Similarity Metric (NSIM) to simulate performance intensity functions in quiet and in noise, which compared favourably with SII predictions of phoneme recognition in NH listeners. Zilany and Bruce (2007) used the model's previous version (Zilany and Bruce, 2006) together with a modified version of the Spectro Temporal Modulation Index (STMI, Elhilali et al., 2003) and found good agreement between the model predictions and SI scores using filtered sentences at different presentation levels and with different levels of cochlear impairment. These two studies have been performed under different settings and conditions (e.g., speech material, noise, even with different versions of the model), making it hard to make a direct comparison of their results.

The objective of this study is to assess SI utilizing different objective measures and to compare their performance. We evaluate them in the same manner and under the same conditions, allowing for an understanding of their possibilities and shortcomings. For the neurogram-based metrics, we use the AN model proposed by Zilany et al. (2009) with parameters defined by Zilany et al. (2014) as a front end to generate neurograms (envelope – ENV, temporal fine structure – TFS, and early stage – ES) at different time scales. As a back end, we use the NSIM and STMI metrics as described by Hines and Harte (2012) and Elhilali et al. (2003), respectively. We compare their performance to two well-established filterbank-based metrics: the SII (ANSI, 1997) and the sEPSM (Jørgensen and Dau, 2011). Finally, we investigate whether these objective measures could predict behavioral scores or not by looking into their correlations with behavioral data. We hypothesize that the neurogram-based metrics will correlate at least as well as the filterbank-based metrics with the behavioral scores, since they encode physiological information that we think is important for speech understanding. We also believe that the neurogram based metrics that have an ENV-based front end will be correlated higher with the behavioral scores than those that do not, since the literature suggests that the ENV component of speech has a large contribution to its perception (Drullman, 1995; Shannon et al., 1995; Smith et al., 2002; Swaminathan and Heinz, 2012).

The chapter is organized as follows. Section 2.3 provides the technical background of the objective measures used in this work. Section 2.4 describes how the study was conducted. Section 2.5 presents the obtained results, which are further discussed in Sec. 2.6. Section 2.7 closes the chapter with our overall conclusions.

2.3 Background

2.3.1 Neurogram-based Metrics

AN Model The model proposed by Zilany et al. (2014, 2009) is capable of reproducing response properties of AN fibers. It is comprised of various modules, each mimicking a particular function of the auditory periphery. First, the stimulus is passed through a filter emulating the middle ear. The output is then passed through a signal path and a control path. The former simulates the behavior of the OHC-controlled filtering properties of the basilar membrane (BM) in the cochlea and the transduction properties of the IHCs by a succession of non-linear and low-pass filters. The latter simulates the function of the OHCs in controlling BM filtering. The output of the control path feeds back into itself and into the signal path, as well. The IHCs output then goes through an IHC-AN synapse module with two power-law adaptation paths, accounting for slow and fast adaptations.

Neurograms For each input, the AN model produces three different neurograms: the ENV, the TFS, and the ES neurograms, which are generally explained as follows. Further details on their implementation are given in Sec. 2.3.1.

The ENV and TFS neurograms[2] allow studying the neural response at different time resolutions (Hines and Harte, 2010, 2012). The ENV neurogram represents smoothed (averaged) discharge rate using a bin size of 6.4 ms. Thus, only slow temporal modulations related to the ENV are available. On the other hand, the TFS neurogram retains spiking information and phase-locking related events (Young, 2008). The TFS neurogram uses a bin size of 0.16 ms. Both of them are obtained by convolving them with a Hamming window of 128 and 32 samples, respectively, with 50 % overlap.

The ES neurogram explicitly encodes temporal envelope modulations due to the interplay of the spectral components in each band (Elhilali et al., 2003). In this case, the neural activity was binned into time bins of 8 ms. It was obtained by convolving it with a rectangular window of 2 samples with a 50 % overlap.

---

[2] The ENV and TFS terminology is not strictly equivalent to the one used by Rosen (1992).

Similarity Metrics

In order to obtain a measure of SI for each speech token, two different neurograms are computed: a reference neurogram r (which receives the speech token in quiet as an input) and a degraded neurogram d (which receives the speech token in noise as an input). Then, the similarity between the two neurograms can be calculated using different metrics, in our case the NSIM and the STMI.

NSIM The NSIM is a simplified version of the Structural Similarity Index (SSIM, Wang et al., 2004). It considers the neurograms as images and quantifies their similarity as a function of their luminance l (comparing the mean values across both images) and their structure s (equivalent to their correlation coefficient), as given by Eq. 2.1 and 2.2:

$$\begin{aligned} \mathrm{NSIM}(r,d) &= l(r,d)\cdot s(r,d) && (2.1)\\ &= \frac{2\mu_r\mu_d + C_1}{\mu_r^2+\mu_d^2+C_1}\cdot\frac{\sigma_{rd}+C_2}{\sigma_r\sigma_d+C_2} && (2.2)\end{aligned}$$

In the latter equation, μ and σ represent the mean intensity and standard deviation, respectively, of their corresponding neurograms, while σ_{rd} represents the covariance between both of them. Each factor contains constant values $C_1 = 0.01L$ and $C_2 = (0.03L)^2$ (where L is the intensity range). Although these have little effect on the metric value itself, they are useful to prevent instabilities at boundary conditions.

STMI The STMI is a measure of speech modulation integrity. It quantifies the degradation of the speech modulations in the temporal and the spectral dimensions jointly due to the addition of noise (regardless of its nature) or processing of the speech signal itself. It takes the AN activity and projects it into a higher, more central level at the primary auditory cortex. To do so, it applies a bank of modulation selective filters (which resemble those described in the mammalian central auditory system, Chi et al., 1999; Wang and Shamma,

1995) to the input neurograms. The former consists of 9 temporal filters ranging from slow to fast rates and of 11 spectral filters ranging from narrow to broad scales. This yields a 4D representation of the activity at a central (i.e., cortical) level: time, frequency, temporal rate, and spectral scale. Then, in order to only extract temporal and spectral modulations, the cortical representation is adjusted by subtracting the ES neurogram of a base signal (i.e., a signal with the same long-term spectrum but randomized phase). Finally, the STMI is calculated between the reference cortical representation R of the neurogram r (corresponding to the word/phoneme in quiet) and the cortical representation D of the neurogram d (corresponding to the word-phoneme in noise) using Eq. 2.3:

$$\mathrm{STMI}(R, D) = 1 - \frac{||R - D||^2}{||R||^2} \tag{2.3}$$

2.3.2 Filterbank-based Metrics

The SII and the sEPSM are two well established filterbank-based SI metrics. The SII is a well known measure of SI that uses a relatively simple filterbank to model the ear's frequency selectivity. The sEPSM also includes such a filterbank, but it goes one step further and uses an additional (modulation) filterbank.

SII The SII computes the average amount of useful speech information that is available to the listener (ANSI, 1997). Mathematically, it is given by Eq. 2.4.

$$\mathrm{SII} = \sum_{i=1}^{n} \mathrm{FIF}_i A_i \tag{2.4}$$

It receives as an input the *clean* speech and the noise signal (Fig. 2.1). First, it partitions the inputs into n individual frequency bands. These can be one-third-octave bands, octave bands, or critical bands. Next, for each band, its audibility A (i.e., the proportion of audible speech cues that are audible to the listener) is calculated. A is simply based on the level of speech relative to the level of noise.

For its computation, the spectrum level of noise is subtracted from the spectrum level of speech in each band. Then, correction factors (designed to account for distortion due to high presentation levels and upward spread of masking) are applied. Lastly, the SNR is computed and normalized between 0 and 1 (assuming a dynamic range of speech of 30 dB). Next, A is multiplied by the band frequency importance function (FIF), which determines the contribution of different frequency regions to speech recognition. The FIF depends on the type of speech material and presentation level. The sum of these values across all bands is approximately equal to 1. Finally, these values are summed across the different frequency bands, yielding a single SII value (Hornsby, 2004).

sEPSM The sEPSM (Jørgensen and Dau, 2011) is an extension of the EPSM. It was originally proposed by Dau et al. (1999) and Ewert and Dau (2000).

It receives as an input the *degraded* speech and the noise signal (Fig. 2.1). First, it passes each input through a gammatone filterbank. Then, the envelope of each channel is extracted using the Hilbert transform. The resulting envelope is input to a modulation filterbank and the power of the filtered envelope computed, resulting in $P_{\mathrm{env}\,S+N}$ and $P_{\mathrm{env}\,N}$ for the degraded speech signal and the noise signal, per channel, respectively. After that, the envelope SNR of a channel i ($\mathrm{SNR}_{\mathrm{env}\,i}$) is computed using Eq. 2.5. Next, the $\mathrm{SNR}_{\mathrm{env}}$ of all n channels is combined into a single overall value, using Eq. 2.6.

$$\mathrm{SNR}_{\mathrm{env}\,i} = \frac{P_{\mathrm{env}\,S+N} - P_{\mathrm{env}\,N}}{P_{\mathrm{env}\,N}} \tag{2.5}$$

$$\mathrm{SNR}_{\mathrm{env}} = \sqrt{\sum_{i=1}^{n} (\mathrm{SNR}_{\mathrm{env}\,i})^2} \tag{2.6}$$

Following, the overall $\mathrm{SNR}_{\mathrm{env}}$ is transformed to a sensitivity index d' of an ideal observer using Eq. 5.2, where k and q are speech-material-dependent parameters.

$$d' = k \cdot \mathrm{SNR}_{\mathrm{env}}^{q} \tag{2.7}$$

Finally, d' is converted into the probability of the ideal observer of correctly recognizing the speech item P_{correct} using the mAFC model proposed by Green and Birdsall (1964). This model compares the input speech element with a set of m previously stored alternatives. Then, it chooses the most similar one, x_S. x_S is a random variable with mean d' and variance σ_S^2 (which is related to the redundancy of the speech material). The remaining $m-1$ items are considered to be noise. Of these, the one that has the largest similarity with the input speech element is chosen as x_N. The latter is also a random variable with mean μ_N and variance σ_N^2. P_{correct} is calculated from the difference distribution of x_S and x_N, as given by Eq. 2.8. Φ stands for the cumulative normal distribution.

$$P_{\text{correct}} = \Phi\left(\frac{d' - \mu_N}{\sqrt{\sigma_S^2 + \sigma_N^2}}\right) \tag{2.8}$$

2.4 Materials & Methods

An overview of the materials and methods used is provided in Fig. 2.1.

2.4.1 Speech Material

The *Leuven Intelligibility Peutertest* (Lilliput) speech material was used in this experiment. It consists of 378 meaningful Flemish CVC words uttered by a female speaker. In order to improve its homogeneity, we selected words that were within one standard deviation around the mean of their average speech recognition threshold (SRT) for adults (i.e., words with an SRT within the -9.8 ± 2.9 dB range). A subset of 65 randomly-picked words was finally chosen.

Stimuli were combined with the accompanying long-term averaged speech-shaped noise at five different SNRs: from 0 to -12 dB in steps of 3 dB. These degraded audio files were the target material for the objective measures (Sec. 2.4.2) and the behavioral measurements (Sec. 2.4.3).

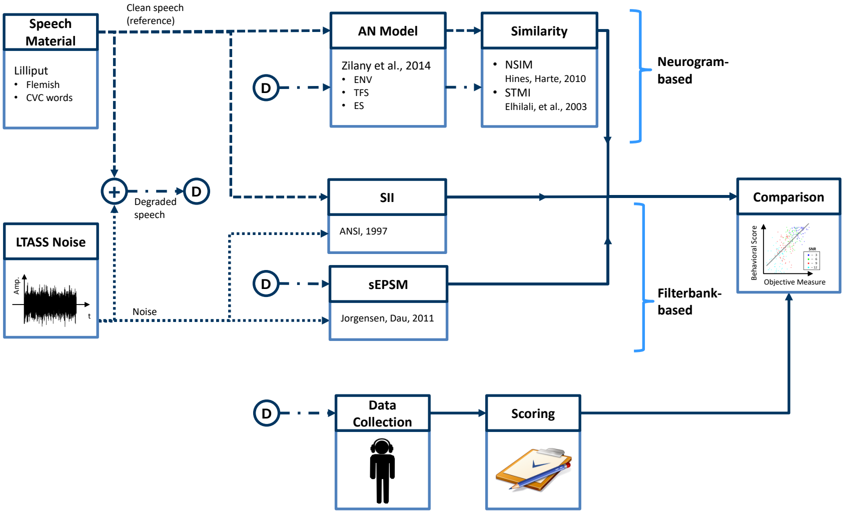

Figure 2.1: Overview of the materials and methods used in this study. Dashed arrows represent the reference (clean) signal. Dotted arrows represent the noise signal. Dashed-and-dotted arrows represent the degraded (clean + noise) signal.

Segmentation

On one hand, we were interested in studying perception at a *word level* (i.e., with no sentence context), since it has been shown that sentence context has an influence on word perception (Boothroyd and Nittrouer, 1988), which cannot be easily modelled. On the other hand, we were interested in studying perception at a *phoneme level*, given that phoneme scores present a reduced variability and thus exhibit greater test-retest reliability (Gelfand, 1998).

Thus, for the latter each audio file was manually segmented into phonemes. Two different kind of segmentations were done. In the first, the segmented audio only included the sound of its corresponding phoneme, yielding segments C_1, V, and C_2 (*pure-phoneme segments*). In the second, the segmented audio additionally included the transition to and/or from the preceding/succeeding phoneme, yielding segments Cv, cVc, and vC (*transitions-included segments*). The phoneme limits were determined using Praat 5.3.16 (Boersma and Weenink, 2014). These were delimited by visual inspection of the time signal and the spectrogram, together with auditory inspection. Figure 2.2 shows an example segmentation of the word *bot*.

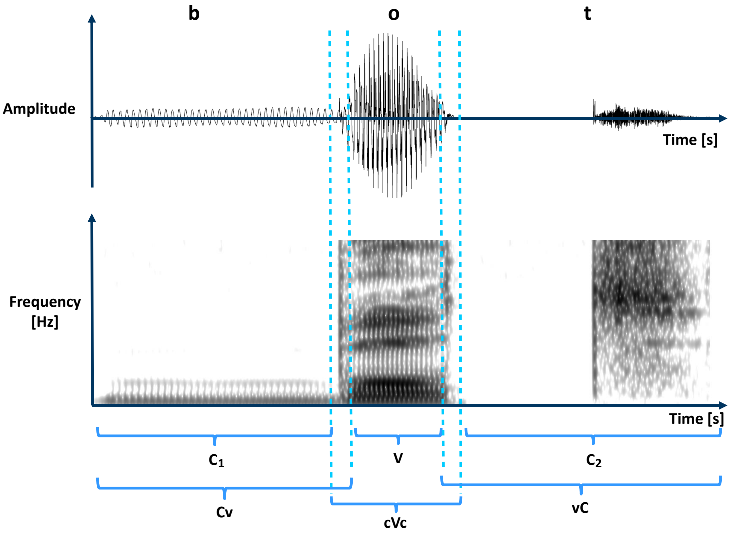

Figure 2.2: Example segmentation of the word *bot*. The time signal is shown in the top part. The spectrogram is shown in the bottom part.

2.4.2 Objective Measures

Neurogram-based Metrics

The degraded and clean (reference) signal were fed to the mathematical model of the AN proposed by Zilany et al. (2014, 2009) with parameters set to simulate NH listeners. Each input produced an ENV, a TFS, and an ES neurogram (Sec. 2.3.1).

All neurograms depicted the average response of 50 AN fibers at each CF with different spontaneous rates: high (100 spikes/s), medium (5 spikes/s), and low (0.1 spikes/s), with weights of 0.6, 0.2, and 0.2, respectively, corresponding to the distribution observed in animals (Zilany and Bruce, 2007). The ENV and TFS neurograms considered 30 CFs logarithmically spaced from 250 to 8000 Hz (Hines and Harte, 2010, 2012). The ES neurogram considered 128 CFs logarithmically spaced from 180 to 7000 Hz (Elhilali et al., 2003). Figure 2.3 shows example ENV and TFS neurograms in quiet and at different SNRs for the word *bot*. Figure 2.4 shows example ES neurograms and cortical representations for the same word under the same conditions.

After that, the deterioration from the reference speech token r to the token of the degraded stimulus d was quantified using the NSIM and the STMI. Specifically, the NSIM metric was applied to the ENV and TFS neurograms (Hines and Harte, 2010, 2012). The STMI metric was applied to the ES neurogram (Elhilali et al., 2003). Additionally, we applied the STMI to the ENV and TFS neurograms, as well, in order to explore the results of projecting the information of such neurograms to a higher (more central) level. Care was taken to make sure that the STMI modulation filters covered the correct range of spectral modulation scales for the corresponding neurogram's CFs. Thus, the STMI spectral filters ranged from 0.25 to 8 cycles/oct, while the STMI ENV and STMI TFS spectral filters ranged from 0.25 to 2 cycles/oct. The temporal filters in both cases went from 2 to 32 Hz. For all metrics, the value for each word/phoneme was averaged across participants for each SNR condition. Lastly, a straight line was fitted through these points.

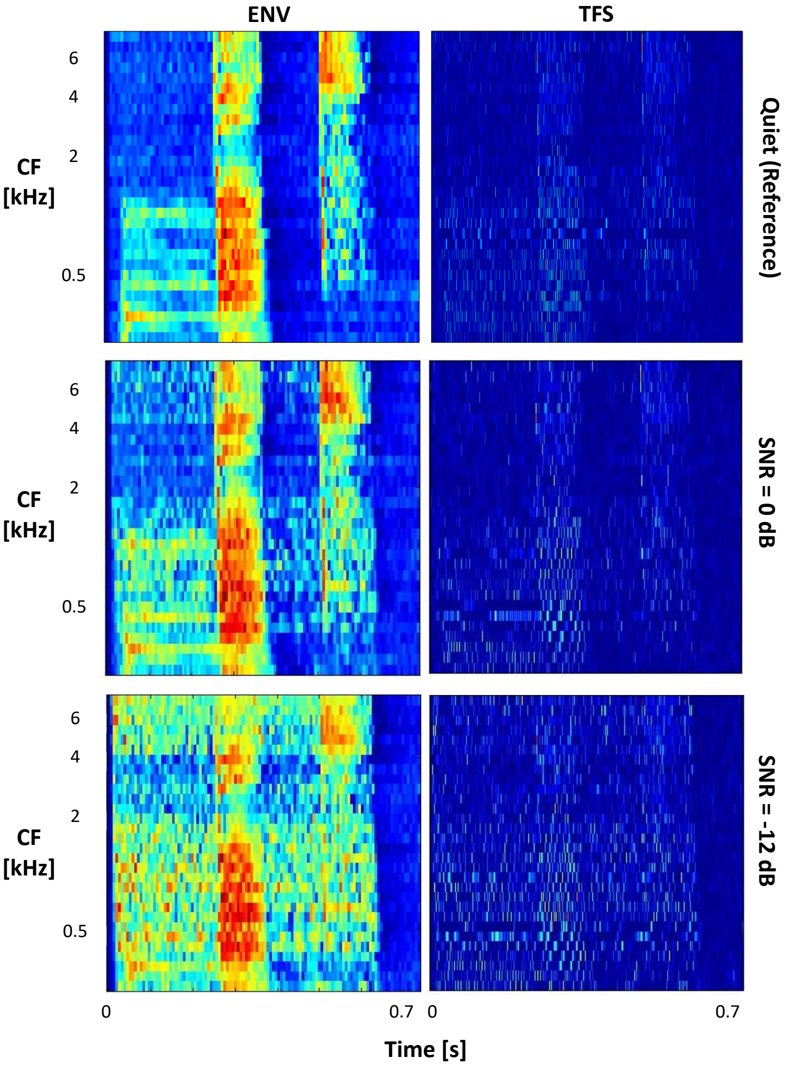

Figure 2.3: Example ENV (left column) and TFS (right column) neurograms for the word *bot*. Top row corresponds to the reference condition (in quiet). Middle row corresponds to the condition SNR = 0 dB. Bottom row corresponds to the condition SNR = −12 dB. Notice how information is represented differently by the ENV and TFS neurograms.

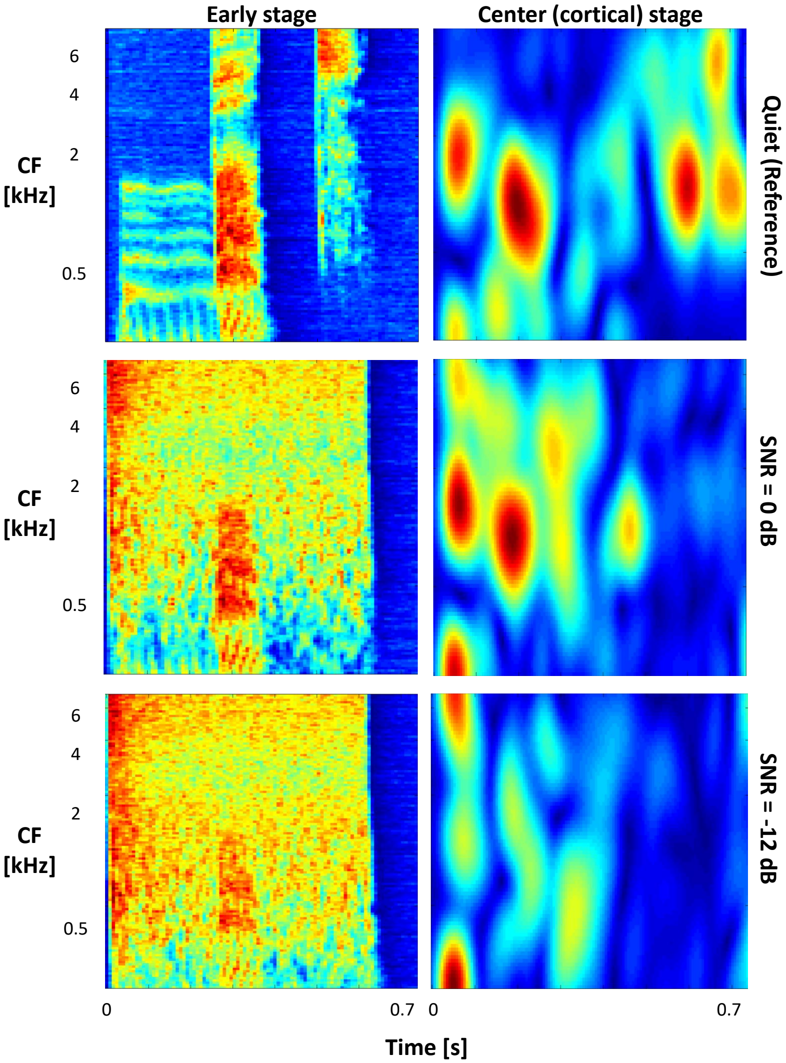

Figure 2.4: Example ES neurograms (left column) and cortical representation (right column) for the word *bot*. Top row corresponds to the reference condition (in quiet). Middle row corresponds to the condition SNR = 0 dB. Bottom row corresponds to the condition SNR = −12 dB. Cortical representations correspond to an example temporal modulation rate of 5.65 Hz and a spectral modulation filter scale of 1 cycle/oct.

Filterbank-based Metrics

In order to be make our results comparable with those of previous studies (e.g., Hines and Harte, 2012; Hossain et al., 2016; Mamun et al., 2015), we chose to partition the input spectra into one-third octave bands for the computation of the SII. Furthermore, we used the SPIN FIF (ANSI, 1997, Table B.2).

In the case of the sEPSM, we used values of 0.275 and 0.315 for the ideal observer parameters k and q, respectively. These were empirically obtained by minimizing the root mean square error (RMSE) between the model predictions and the psychometric function of the behavioral data. In order to reduce overfitting, we obtained these values using only one-third of the available data. To convert d' into P_{correct}, m was assumed to be 8,000 (the size of a person's active vocabulary, Müsch and Buus, 2001), since we used an open set.

2.4.3 Behavioral Measurements

Data Collection

Twenty participants (5 males, 15 females, mean age 20.75 ± 1.48 years old) volunteered for this study. They were tested and confirmed to have NH. All of them were native Flemish speakers and provided written informed consent. The study was approved by the local ethics committee.

Measurements were performed in a sound booth. Words were presented to the participant at a fixed speech level of 65 dB sound pressure level (SPL) using APEX 3 (Francart et al., 2008). These were routed from a computer via an external RME Fireface sound card to Sennheiser HD 250 Linear II headphones. Words were presented randomly across SNRs. For each trial, participants were instructed to listen to the stimulus and to type the word they thought they had heard into the computer. Each word was presented once for each SNR condition.

Scoring

For each trial, two different behavioral scores were obtained: a *phoneme score* (at a word level) and a *per-phoneme score* (at an individual phoneme level).

The phoneme score was assigned by the experimenter to the participant's answer depending on the number of phonemes that were correct compared to the original word. For example, if the presented word was *bot* and the participant's answer was *bol*, a phoneme score of 2 was given. Phoneme scores for each word and SNR were averaged across participants.

Per-phoneme (i.e., individual phoneme) scores can be obtained in two ways. On one hand, they can be computed *a posteriori* by comparing the participant's response to the original word and assigning a 1 or a 0, indicating if each phoneme was correct or not, respectively. On the other hand, the phonemes can be presented individually to the participants in a controlled context (e.g., aCa or pVp). For this study, we preferred the former approach, since in this case the phonemes are presented in a more natural context and in a more similar to way to real-life realizations. Although there might be some lexical influence (Ganong, 1980), given the token size we expect that most (if not all) contextual effects are present, thus limiting certain lexical biases in the responses. Furthermore, the speech material used here consists of very short words (CVC), which minimizes such effects even more. Following the previous example, if the presented word was *bot* and the participant's answer was *bol*, a per-phoneme score of [1 1 0] was calculated. Automatic computation of the per-phoneme scores was done using the algorithm proposed by Francart, Moonen and Wouters (2009). The algorithm's performance was evaluated by comparing it with the annotated phoneme score (in the end, the phoneme score is the sum of the three per-phoneme scores), with success in 94% of the cases. The rest of the instances were evaluated manually. Per-phoneme scores were also averaged across participants.

2.4.4 Comparison

We observed a ceiling effect in the behavioral scores, particularly at the condition of SNR = 0 dB. Since this cannot be modelled by most of our objective measures, these data points were not considered for further analysis.

Scatter plots of the scores versus the metrics across different SNRs were made at word and phoneme level. We computed a simple linear regression model in each case. Pearson correlation coefficients were computed between the behavioral and the objective variables. For their comparison, we used William's test (Williams and Williams, 1959) with Bonferroni correction. Additionally, the goodness of the linear regression was evaluated using the F-ratio, which quantifies the improvement of the model prediction compared to the level of the model inaccuracy (Field et al., 2012).

Finally, we evaluated what unique proportion of the variance was explained by the different objective measures. Using a hierarchical predictor selection approach, we chose the metrics of each of the objective measures groups (neurogram-based and filterbank-based) with the highest correlations and used these to generate multivariate linear regression models. Then, we computed the R^2 values for each case and looked at the difference between these values and those obtained in the simple linear regression.

2.5 Results

The results for the objective measures at word level are shown in Fig. 2.5. It depicts the boxplot of these metrics, as well as its average across the whole set of 65 words. The dotted line corresponds to the fitted straight line. Overall, we found a directly proportional relation between the objective metrics and SNR at a word and phoneme level.

Figure 2.6 shows the scatter plots of phoneme scores vs different metrics averaged across participants at a word level. In this case, each point corresponds to a word. For each metric, the Pearson correlation was calculated between the phoneme score and the objective measure and is shown on the bottom right corner in each plot, together with its Bonferroni corrected p-value. ENV-based metrics had a stronger correlation with behavioral scores than TFS-based metrics: the NSIM ENV correlation (0.73) was significantly higher ($p < 0.01$) than the NSIM TFS correlation (0.67); the STMI ENV presented a 0.69 correlation, which is significantly higher ($p < 0.001$) to its TFS counterpart (0.24). In the case of the filterbank-based measures, the SII and sEPSM metrics had a correlation of 0.61 and 0.35, respectively, with $p < 0.001$ for both of them. Overall, the NSIM ENV and STMI ENV showed the strongest correlations of all the metrics ($p < 0.05$) with no significant differences between them ($p > 0.05$). Additionally, we computed the F-ratios for the linear regression of each of the different objective measure (also shown in their corresponding scatter plot together with its p-value). Similar to the Pearson correlations, the NSIM ENV, NSIM TFS, and STMI ENV metrics had the largest F-ratios (275, 225.9, and 219.9, respectively, all with $p < 0.001$). In the case of the filterbank-based measures, the SII and sEPSM metrics had F-ratios of 120.6 and 28.4, respectively, with $p < 0.001$ for both of them.

Based on these results, we chose the NSIM ENV and STMI ENV metrics from the neurogram-based group and the SII from the filterbank-based group as predictors for the multivariate linear regression[3]. Analysing the R^2 values allowed us to quantify the amount of unique proportion of variance explained by the different objective measures. For instance, the regression model that started with the SII had an R^2 value of only 0.37. When we incorporated either the NSIM ENV or STMI ENV metrics into it, the R^2 value significantly increased by 0.18 (ANOVA, $F(1, 259) = 81.51$, $p < 0.001$) and 0.16 (ANOVA, $F(1, 259) = 70.95$, $p < 0.001$), respectively. Going the other way around, when the regression model started with the NSIM ENV or STMI ENV metrics, the R^2 values were of 0.54 and 0.47, respectively (which are already higher than in the SII model case). When we incorporated the SII metric, the R^2 values significantly increased by barely 0.01 (ANOVA, $F(1, 259) = 4.48$, $p < 0.05$) and 0.06 (ANOVA, $F(1, 259) = 25.38$, $p < 0.001$), respectively. A summary of these results is shown in in Table 2.1.

Table 2.1: Results of the (multivariate) linear regression models

Predictors (objective metrics)			R^2	F-statistic	DoF	p-value
NSIM ENV	STMI ENV	SII				
✓			0.54	199.40	260	< 0.001
	✓		0.47	146.10	260	< 0.001
		✓	0.37	86.22	260	< 0.001
✓		✓	0.55	103.90	260	< 0.001
	✓	✓	0.53	96.02	260	< 0.001

---

[3] We checked for collinearity by performing a variance inflation factor analysis. In all cases, it yielded values <10, which indicate a low level of collinearity and therefore little impact on the regression model (Field et al., 2012).

Additionally, Fig. 2.7 shows the Pearson correlation between the per-phoneme scores vs different metrics for pure phoneme segments and for phoneme segments that include transitions. In this case, weak to moderate correlations were mostly found. We were interested in studying the influence of including transitions on the correlation between the objective measures and the behavioral per-phoneme scores. We found significantly lower correlations of C_1, V, and C_2 compared to their counterparts Cv, cVc, and vC in the NSIM ENV ($p < 0.05$, $p < 0.001$, and $p < 0.05$, respectively); significantly lower correlation of C_2 compared to vC in the STMI TFS ($p < 0.05$); significantly lower correlation of C_1 compared to Cv in the SII ($p < 0.05$). The rest of the metrics did not show significant differences.

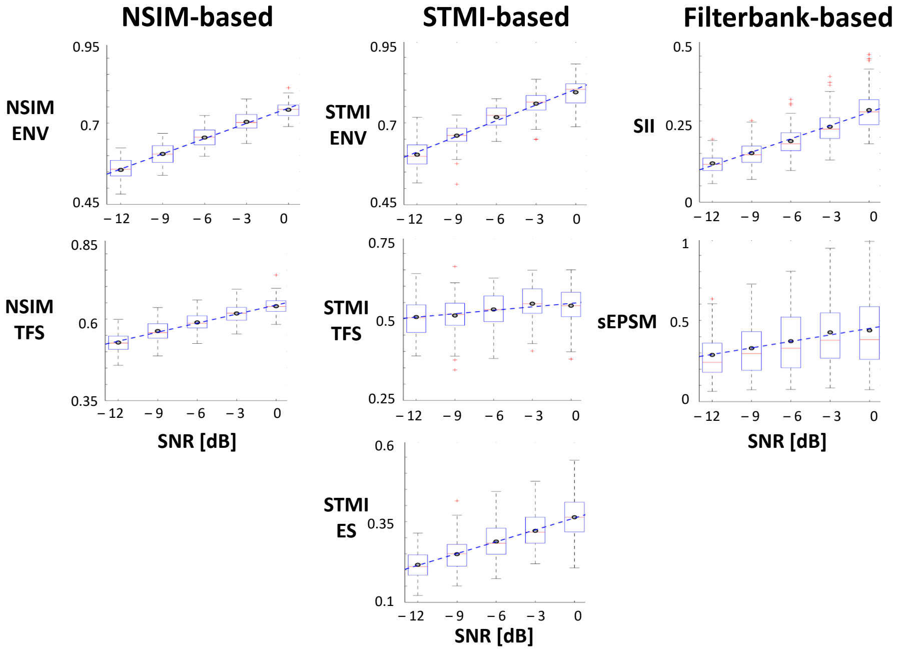

Figure 2.5: Distribution of different objective measures of the complete set averaged across words. Circles represent the mean. The dotted line represents the fitted straight line. Crosses represent outliers. Note that although all objective measure have the same range (0 to 1), they are of a different nature.

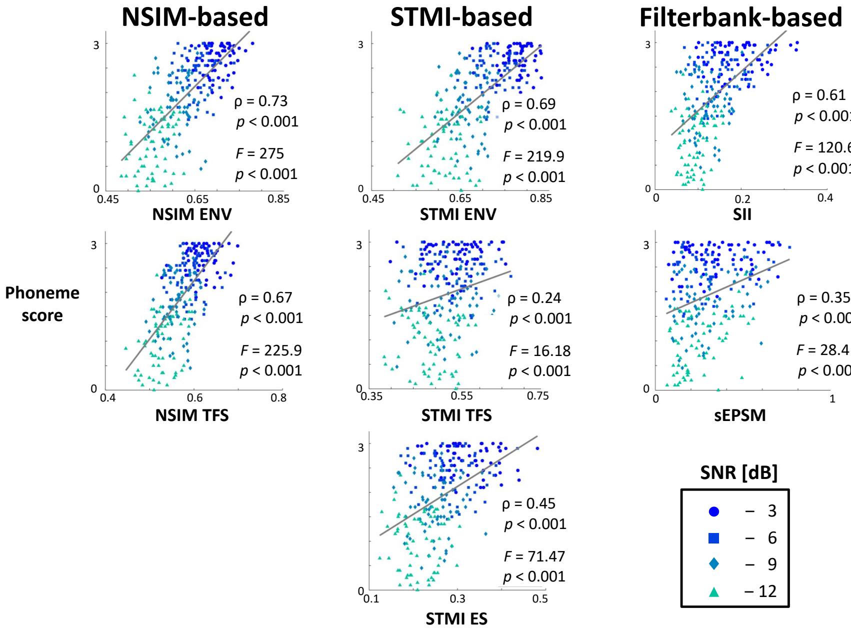

Figure 2.6: Scatter plots of phoneme scores vs different objective measures averaged across participants at a word level (thus each point is a word). Reported correlations and F-ratios were calculated with $\alpha = 0.05$ and significant.

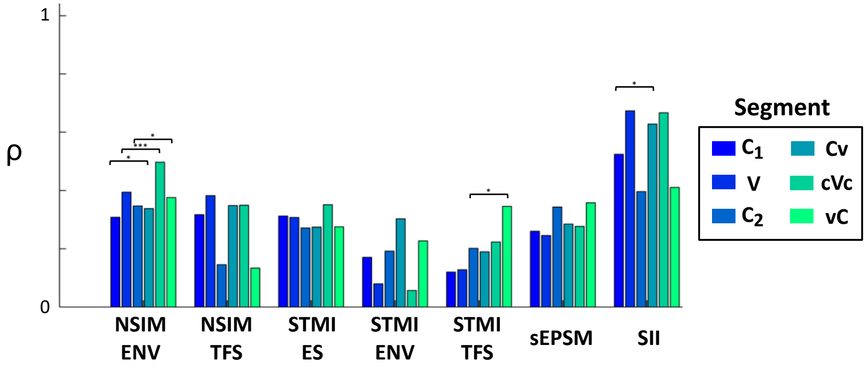

Figure 2.7: Correlation between the behavioral scores and the computed metrics (per phoneme). Reported correlations were calculated with $\alpha = 0.05$.

2.6 Discussion

In this study, we assessed SI using different neurogram-based and filterbank-based objective measures. Then, we calculated the correlations between these and behavioral scores of NH listeners.

Figure 2.5 shows how metric values increase together with SNR. This is due to the fact that noise is random and spurious information which shows in the neurograms as activity (Fig. 2.3), but that actually diminishes the metric.

At a word level, the strongest correlations were found between the behavioral scores and the ENV-based metrics NSIM ENV and STMI ENV, showing correlations of 0.73 and 0.69, respectively (with no significant difference between them, $p > 0.05$). These are significantly higher than their TFS counterparts: NSIM TFS with a value of 0.67 and STMI TFS with a value of 0.24 ($p < 0.05$ and $p < 0.001$, respectively); they were also significantly higher than those found using filterbank-based metrics: SII with a value of 0.61 and sEPSM with a value of 0.35 ($p < 0.01$ and $p < 0.001$, respectively). This suggests that the ENV component of speech (as represented by the ENV neurograms) has a stronger correlation with the behavioral scores than the TFS component. This goes in line with what has been reported in literature: TFS may be important for speech understanding in some conditions (e.g., in settings with background noise, Lorenzi et al., 2006), but generally the ENV component of speech has a much larger contribution to speech perception (e.g., Drullman, 1995; Shannon et al., 1995; Smith et al., 2002; Swaminathan and Heinz, 2012).

Furthermore, it hints that neurogram-based metrics correlate as much (or higher) than filterbank-based metrics. This could be due to the fact that the former incorporate physiological aspects of the auditory system as part of the AN model. The STMI TFS metric had the lowest correlation of all (0.24), while also showing large variance across different listeners. This suggests that the forenamed objective measure is not a reliable predictor for SI. We hypothesize this is because TFS is lost in the base signal of the STMI. Furthermore, the information of the TFS is scattered and sparse and thus incapable of fully reflecting the modulations of the original speech token. These modulations are the base of the cortical representation of the STMI metric, hence it fails to provide a representation of the original input. The computed F-ratios show that of the proposed objective measures, the linear regression models of the

NSIM ENV ($F = 275$, $p < 0.001$), NSIM TFS ($F = 225.9$, $p < 0.001$), and STMI ENV ($F = 219.9$, $p < 0.001$) metrics fit the data the best, since their large values reflect smaller differences between the model's predictions and the observed data. On the contrary, the STMI TFS model ($F = 16.18$, $p < 0.001$) had the smallest F-ratio, reinforcing the idea that in the presented framework, this metric is a poor predictor of SI.

The simple linear regressions showed that from the NSIM ENV, STMI ENV (both neurogram-based), and the SII (filterbank-based) metrics, the latter had the smallest R^2 value. Furthermore, the R^2 values of the multivariate linear regressions showed the smallest improvement when the SII was incorporated. This suggests that (ENV) neurogram-based metrics are able to account for a larger proportion of the variance than filterbank-based metrics (SII, in this case), endorsing their value for SI prediction in the presented framework.

At a phoneme level, we found significantly lower correlations of C_1, V, and C_2 compared to their counterparts Cv ($p < 0.05$), cVc ($p < 0.001$), and vC ($p < 0.05$) in the NSIM ENV case. The fact that this was the only metric that was consistently sensitive to phoneme transitions could suggest that that these have a larger impact on intelligibility on the ENV component of speech (where they can be captured), rather than in the TFS. Furthermore, this could also hint that phoneme transitions have a larger impact on the information at the AN level, rather than at a higher level (e.g., cortical representation). However, these results have to be handled with care, since there is no clear agreement in the literature regarding the impact of transitions in different speech intelligibility tasks. On one hand, it has been suggested that transitions have an important impact on phoneme identification. Jenkins et al. (1994) found that the onset and offset of vowels in /dVd/ syllables was enough to identify the syllable. Strange and Bohn (1998) showed that perceptual differentiation of German vowels is dependent on spectral information contained in transition cues (e.g., onsets and offsets). On the other hand, the opposite hypothesis has also been proposed. Cole et al. (1996) found that the location of the segment boundaries (i.e., the inclusion or exclusion of phoneme transitions) does not have a strong impact in consonants or vowels intelligibility. Using an entropy-based approach, Stilp and Kluender (2010) found that intelligibility patterns of replaced vowel-consonant and consonant-vowel transitions were indistinguishable from that of vowels. Fogerty and Kewley-Port (2007, 2009) concluded that for speech tokens with vowels present, the information in the transitions does not contribute to intelligibility since it might be redundant with that found at the center of the vowel (they agreed, though, that for consonant-only speech tokens, information in the transitions did provide a perceptual benefit). Lee and Kewley-Port (2009) found that different transitional information had a similar effect in speech intelligibility for young NH and elderly HI listeners.

Although comparing the metrics' performance between the word and phoneme level cases was not the main objective of this analysis, it is still worth mentioning a few things. Overall, we found stronger correlations in the former compared to the latter case. We hypothesize this could be due to a variety of reasons, depending on the metric. For the NSIM-based metrics, we think that using shorter neurograms reduces the amount of useful information available for the metric when comparing the neurograms. In the case of the STMI-based metrics, we believe that the length of the phoneme segments is not long enough to be captured properly by all the filters of the temporal filter bank.

Finally, even though physiological-inspired frameworks (such as the one presented here) are successful in predicting SI, they still have a few shortcomings that are worth pointing out. Speech perception is a very complex process. The mapping of the speech signal along the auditory pathway is an intricate mechanism that is not yet fully understood. Studying SI using different approaches (Allen, 2005) is the first step towards a better comprehension of the various processes involved (e.g., the study of the learning component of speech understanding has served as the base of the development of automatic speech recognition systems, Benzeghiba et al., 2007). Additionally, current biologically-inspired models have been validated mostly using animal data. Their translation to human auditory processes rely on several assumptions, many of which still need to be confirmed by further physiological studies.

2.7 Conclusions

In this work, we investigated the correlation of different objective metrics with behavioral scores, with special emphasis on neurogram-based metrics that use the AN model proposed by Zilany et al. (2014, 2009) as a front end.

The relation between the objective measures (when averaged across participants and across words) and SNR can be explained by fitting a straight line. Furthermore, we found significantly stronger correlations between behavioral measurements and the ENV based objective measures at a word level. This goes in line with the usefulness of the ENV for behavioral perception, with the NSIM ENV and STMI ENV presenting the strongest ones and being able to explain the largest variance proportion. Besides, these objective measures present a few more advantages over the rest. Since they are based on the responses of the AN, they inherently incorporate physiological information. This provides a more transparent approach to understanding the processes occurring in the auditory system. Furthermore, thanks to the versatility of the AN model, it allows to incorporate biologically the effects of hearing loss due to damage to the IHCs, OHCs or both, something that is not possible to do straightforwardly using one of the filterbank-based approaches. Additionally, the latter rely heavily on calibration of their parameters for different cases (e.g., speech material, noise conditions), which is not only hard to achieve, but it can also lead to overfitting. At a per-phoneme level, we found that the NSIM ENV was consistently sensitive to the phoneme transitions for C_1, V, and C_2. Lastly, we could not find evidence that simulating processes at a central level using the current approach (i.e., applying a cortical model on top of a peripheral representation) provides extra benefit over the information already available at the AN.

CHAPTER 3

The Effect of Level on Spectrotemporal Modulation Detection[1]

---

[1] The work presented in this section is currently in press as Moncada-Torres, A., Magits, S., Van Deun, L., Wouters, J., van Wieringen, A., and Francart, T. (**2018**). "The Effect of Presentation Level on Spectrotemporal Modulation Detection," Hearing Research. Changes are limited to layout, graphical appearance, and minor editing.

3.1 Abstract

Speech understanding in noise relies (at least partially) on spectrotemporal modulation sensitivity. This sensitivity can be measured by different spectral ripple tests, which can be administered at different presentation levels. However, it is not known how presentation level affects spectrotemporal modulation thresholds. In this work, we present behavioral data for normal-hearing adults which show that at higher ripple densities (2 and 4 ripples/oct), increasing presentation level led to worse discrimination thresholds. Results of a computational model suggest that the higher thresholds could be explained by a worsening of the spectrotemporal representation in the auditory nerve due to neural activity saturation. Our results demonstrate the importance of taking presentation level into account when administering spectrotemporal modulation detection tests.

3.2 Introduction

Complex acoustic signals such as speech are characterized by a combination of spectral and temporal modulations. Speech understanding relies (at least partially) on the ability to detect and discriminate these modulations. In other words, it relies on an individual's spectrotemporal modulation sensitivity (Supin et al., 1997). This can be assessed by two categories of tests: *spectral ripple discrimination* tests and *spectral/spectrotemporal modulation detection* (SMD/STMD, respectively) tests. There are many varieties of these. However, in this paper we focus on an SMD/STMD test where participants are asked to discriminate between a modulated and unmodulated stimulus. The modulation detection threshold is usually defined as the minimal peak-to-valley ratio or modulation index at which the participant can discriminate between the two stimuli (e.g., Bernstein et al., 2013).

It has been shown that SMD/STMD thresholds are correlated with different measures of speech perception in quiet and in noise (Anderson et al., 2012; Croghan and Smith, 2018; Davies-Venn et al., 2015; Mehraei et al., 2014). Additionally, they can provide a non-linguistic measure of spectral/spectrotemporal sensitivity without the confounding factor of language knowledge that plays a role in standardized tests (e.g., speech audiometry, Choi et al., 2016; Davies-Venn et al., 2015; Gifford et al., 2014). This has motivated their use for a variety of purposes. For example, STMD paradigms have been used to explore the perceptual learning mechanisms in the auditory system (Sabin et al., 2012). SMD/STMD tests have also used spectral/spectrotemporal resolution successfully as an outcome measure in different fields of audiological research: prediction of speech understanding in noise of hearing-aid users (Bernstein et al., 2016), assessment of cochlear implant candidacy, parameter fitting, and new sound processing strategies (Choi et al., 2016; Croghan and Smith, 2018; Langner et al., 2017; Zheng et al., 2017), evaluation of bimodal hearing benefit (Zhang et al., 2013), and music perception (Choi et al., 2018).

Although these tests are used mostly in audiological research, to our knowledge no studies have evaluated how presentation level affects SMD/STMD thresholds. This is relevant because SMD/STMD thresholds might be negatively affected by the broadening of the auditory filters caused by increasing presentation level (Glasberg and Moore, 2000). Taking the effect of level into account is crucial when administering SMD/STMD tests in a research environment, in (potential) clinical practice, and even more in test situations where it cannot be controlled strictly (e.g., home-based computerized rehabilitation programs). Furthermore, we need to understand this effect to be able to make a fair comparison of behavioral SMD/STMD results obtained at different presentation levels within and across studies.

The goal of this work was to explore how presentation level affects SMD/STMD thresholds for young adult NH participants. Specifically, we focused on the STMD test, since spectrotemporally modulated (i.e., moving spectral ripple) stimuli have been suggested to provide a better representation of speech (Won et al., 2015) than stimuli measuring sensitivity to only spectral (i.e., rippled, Litvak et al., 2007; Saoji et al., 2009) or temporal modulation. Additionally, STMD tests allow to prevent participants from having access to phase cues by using low rate temporal modulation (Bernstein et al., 2013). Furthermore, we used a biologically inspired model of peripheral processing up to the auditory nerve (AN) to help us interpret the behavioral results, to study the contribution of peripheral information to spectrotemporal sensitivity, and to generate STMD thresholds predictions.

3.3 Behavioral Measurements

3.3.1 Materials & Methods

Participants Ten participants (1 male, 9 female, median age 23.5 years, age range 21–29 years) took part. They were confirmed to have audiometric thresholds $\leq$ 20 dB HL at all octave frequencies from 125 to 8000 Hz. Written informed consent was obtained. The study was approved by the Ethics Committee of the University Hospitals Leuven (approval no. B322201731501).

Equipment Measurements were performed in a double-walled sound-attenuating booth. Stimuli were played from a computer via an RME Hammerfall DSP Multiface II sound card and presented to the participants through Sennheiser HDA 200 headphones using APEX 3 (Francart et al., 2008).

Stimuli We used the spectrotemporally modulated stimuli described by Kowalski et al. (1996) and Chi et al. (1999). These were 500 ms long (including 20 ms onset and offset cosine ramps) and were generated with a sampling frequency of 44100 Hz and a 16-bit resolution. The spectral modulation was achieved as follows. The spectrum of the ripple stimulus (the "carrier") consisted of 4000 random-phase tones equally spaced along the (logarithmic) frequency axis from 354 to 5656 Hz. The amplitudes of the individual components were adjusted to form a sinusoidally shaped spectrum around a flat base. The amplitude of the ripple was defined as the modulation depth m. The initial phase of the ripple Φ was defined relative to a sine wave starting at the low-frequency edge. Its value was set using 50 different selections of random phases between 0 and 2π to prevent participants from using phase differences as a cue. The ripple density was defined as Ω (with values of 0.5, 2, and 4 ripples/oct). The mathematical expression for the static ripple is given in Eq. 3.1

$$S(x) = 1 + m \sin(2\pi\, \Omega\, x + \Phi) \tag{3.1}$$

where x is the position on the logarithmic frequency axis (in octaves), which was defined as $x = \log_2(\frac{f}{f_0})$ with f being the component tone frequency and f_0 the low-frequency edge. Notice that when $m = 0$, the resulting profile is a flat spectrum.

The temporal modulation was achieved by moving the static ripple downwards along the frequency axis at a constant velocity ω (defined as the number of ripple cycles per second passing the low-frequency edge of the spectrum). In our case, ω had a value of 4 Hz. The complete mathematical expression for the spectrotemporal modulated stimuli is given in Eq. 3.2, where t is time.

$$S(x,t) = 1 + m \sin(2\pi\, (\omega\, t + \Omega\, x) + \Phi) \tag{3.2}$$

In order to make our results comparable to those of previous studies, we report m as $20\ \log_{10}(m)$ (i.e., in dB). The *reference stimulus* was unmodulated (i.e., $20\ \log_{10}(m) = -\infty$ dB), whereas the modulation depth of the *target stimulus* was varied adaptively (Sec. 3.3.1). Figure 3.1 shows spectrograms of the reference stimulus and of two example target stimuli ($20\ \log_{10}(m) = -6$ dB and $20\ \log_{10}(m) = 0$ dB).

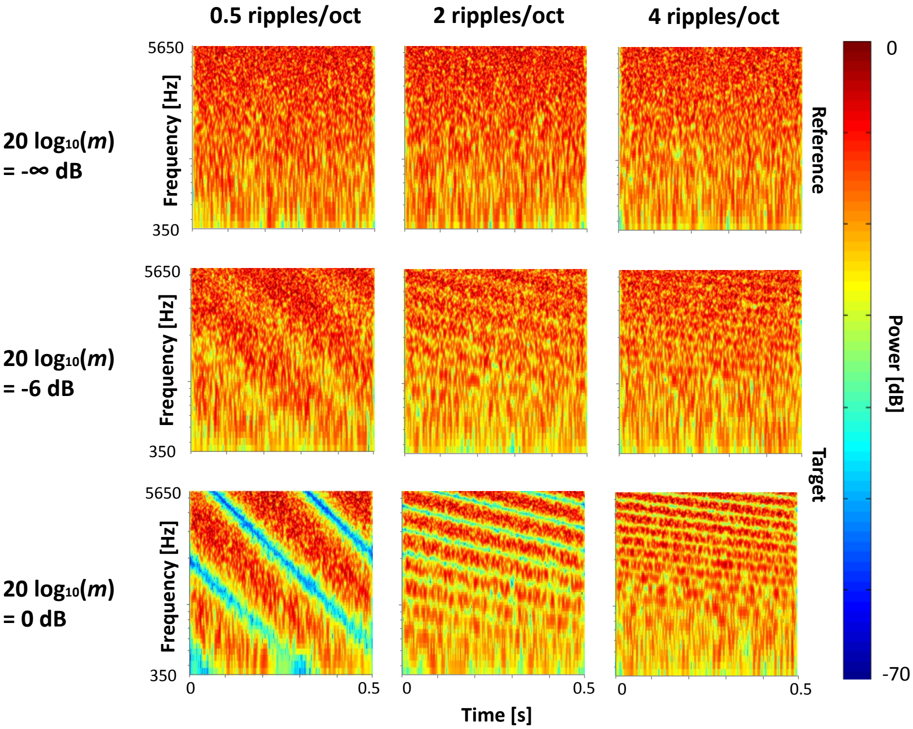

Figure 3.1: Spectrograms of the spectrotemporal modulated stimuli. The pattern along the logarithmic frequency axis changes with ripple density.

Procedure

In a first part, we presented the stimuli at levels of 65 and 86 dB SPL using all three ripple densities (0.5, 2, and 4 ripples/oct). In a second part, we presented stimuli at levels of 55, 65, 75, and 86 dB SPL with a ripple density of 4 ripples/oct. In both parts, stimuli were presented monaurally to the left ear. We used level roving of 8 dB (i.e., random gain between -4 and 4 dB for each stimulus) to reduce the salience of level cues (Eddins and Bero, 2007).

A two-interval two-alternative forced-choice task was used. One of the intervals contained the unmodulated (i.e., reference) stimulus and the other interval contained the modulated (i.e., target) stimulus. The target was randomly presented in the first or second interval with equal probability. There was a 500 ms pause between intervals. Participants were seated in front of a computer screen. They were instructed to discriminate the target interval, which would correspond to the stimulus with a “rippled, vibrating sound”, from the reference interval, which would correspond to the stimulus with a “noisy sound”. They did so by clicking on the corresponding button on the screen (or by using the corresponding keys on the keyboard). Visual feedback was provided through a green (correct response) or red (incorrect response) highlight after each trial. Conditions were presented to each participant in a random order. In a given run, the ripple density was fixed. The modulation depth at threshold was estimated using a three-down one-up procedure tracking the 79.4% point on the psychometric function (Levitt, 1971). Each run started with a fully modulated target ($20 \log_{10}(m) = 0$ dB). The modulation depth was decreased by 6 dB after the first reversal, changed by 4 dB until two more reversals occurred, and changed by 2 dB for the last 6 reversals. A run was ended after 9 reversals. For each run, the mean value of $20 \log_{10} m$ at the last 6 reversals was calculated. Participants completed a test and retest run for every condition. If the thresholds for the two differed by more than 3 dB, a third run was completed. For each condition, the final threshold was taken as the average of all runs.

3.3.2 Results

Figure 3.2 shows the boxplot of the STMD thresholds together with the average across participants for part 1 (65 and 86 dB SPL and 0.5, 2 and 4 ripples/oct). Lower (more negative) thresholds indicate better performance. A general linear model (GLM) showed that ripple density had a significant effect on the STMD thresholds ($\chi^2(1) = 8.26$, $p < 0.001$) as did level ($\chi^2(1) = 11.76$, $p < 0.001$). Moreover, there was a significant interaction effect of ripple density and level ($\chi^2(1) = 24.17$, $p < 0.001$). Tukey *post hoc* tests on the GLM revealed increased thresholds with increasing ripple density at 86 dB SPL, between 0.5 and 4 ripples/oct ($z = 5.83$, $p < 0.001$, confidence interval (CI) [3.64, 8.02]) and between 2 ripples/oct and 4 ripples/oct ($z = 4.82$, $p < 0.001$, CI [2.63, 7.01]). In contrast, thresholds decreased with increasing ripple density at 65 dB SPL between 0.5 and 2 ripples/oct ($z = -3.57$, $p < 0.001$, CI [-5.76, -1.38]) and then increased between 2 and 4 ripples/oct ($z = 2.54$, $p = 0.012$, CI [0.35, 4.73]). The STMD thresholds were significantly lower at 65 dB SPL than at 86 dB SPL at 2 ripples/oct ($z = 3.12$, $p < 0.001$, CI [0.93, 5.31]) and at 4 ripples/oct ($z = 5.40$, $p < 0.001$, CI [3.21, 7.59]), but not at 0.5 ripple/oct ($z = -1.45$, $p = 0.40$, CI [-3.64, 0.73]).

Figure 3.3 shows the boxplot of the STMD thresholds together with the average across participants for part 2 (55, 65, 75, and 86 dB SPL at 4 ripples/oct). STMD thresholds were significantly higher (worse) with increasing level (Friedman's ANOVA, $\chi^2_F(3) = 24.36$, $p < 0.001$). There was a large increase between 65 and 75 dB SPL. *Post hoc* Conover's tests with Holm correction for multiple comparisons revealed significant differences between 55 and 75 dB SPL ($p < 0.001$), 55 and 86 dB SPL ($p < 0.001$), 65 and 75 dB SPL ($p < 0.001$), 65 and 86 dB SPL ($p < 0.001$), and 75 dB SPL and 86 dB SPL ($p < 0.001$). There was no significant difference between the two lowest levels (55 and 65 dB SPL, $p > 0.05$).

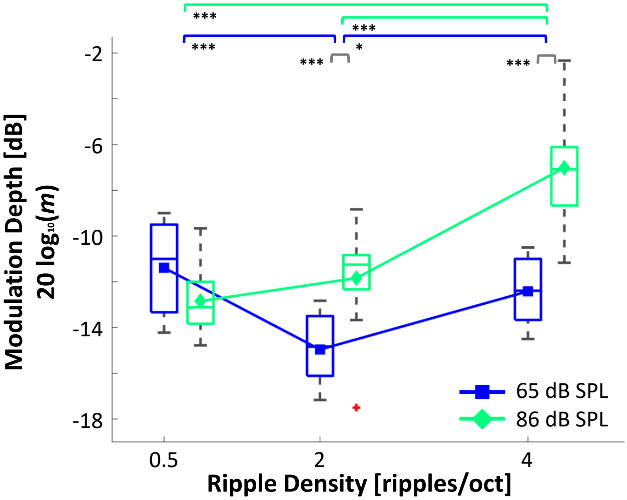

Figure 3.2: STMD thresholds. Lower thresholds indicate better performance. Markers represent the average across participants for each condition. Crosses represent outliers. * $= p < 0.05$, *** $= p < 0.001$.

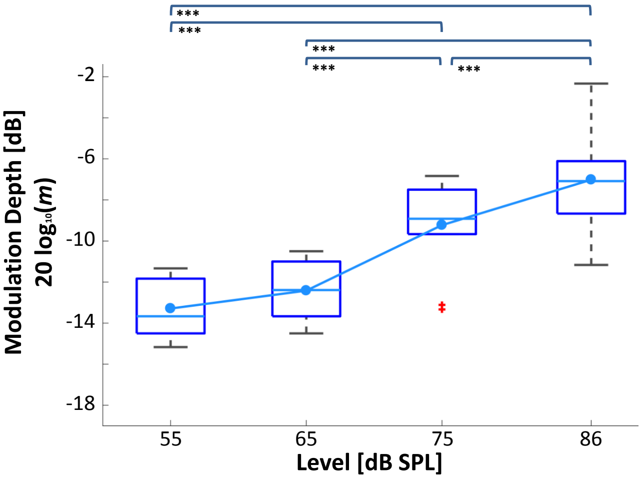

Figure 3.3: STMD thresholds with a ripple density of 4 ripples/oct. Lower thresholds indicate better performance. Markers represent the average across participants for each condition. Crosses represent outliers. *** $= p < 0.001$.

3.4 Computational Model

We used a computational model with a physiologically-inspired front end (i.e., model of the auditory periphery up to the AN) to help us interpret the behavioral results, to study the contribution of peripheral information to spectrotemporal sensitivity, and to obtain quantitative predictions of the behavioral thresholds. Its block diagram is shown in Fig. 3.4. We hypothesized that the model would reflect a detriment in the spectrotemporal representation in the AN with increasing level.

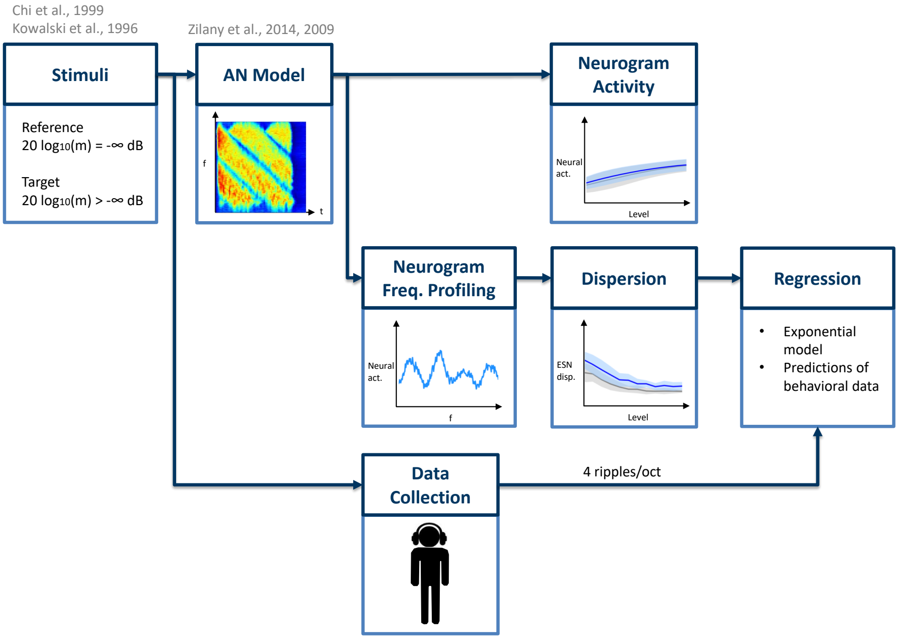

Figure 3.4: Block diagram of the computational model used to interpret the behavioral data and to study the contribution of peripheral information to spectrotemporal sensitivity.

3.4.1 Stimuli

We included a wide range of levels (from 40 to 95 dB SPL in steps of 5 dB). We used the same ripple densities (0.5, 2, and 4 ripples/oct). We simulated responses to the reference stimulus ($20\ \log_{10}(m)\ = -\infty$ dB) and target stimuli with a modulation depth of $20\ \log_{10}(m)\ = -6$ dB and $20\ \log_{10}(m)\ = 0$ dB.

3.4.2 AN Model

The model proposed by Zilany et al. (2014, 2009) was used as a front end. This model reproduces the response of AN fibers to acoustic stimulation. It has been validated with a wide range of physiological data. It is comprised of different modules (each simulating a specific function of the auditory periphery).

First, the stimulus is passed through a filter simulating the middle ear frequency response. The output is fed to a signal path and a control path. The signal path mimics the behavior of the outer-hair-cell- (OHC-) controlled filtering of the basilar membrane in the cochlea and the transduction of the inner-hair-cells (IHCs) by a series of non-linear and low-pass filters. The control path mimics the function of the OHCs in controlling basilar membrane filtering. The control path output feeds back into itself and into the signal path. The output of the IHCs is fed to the IHC-AN synapse module with two power-law adaptation paths, which simulate slow and fast adaptation.

For each stimulus, the AN model generated a so-called early stage neurogram (ESN). An ESN is a time-frequency representation of a signal which encodes temporal modulations caused by the interaction of spectral components in each band (Elhilali et al., 2003). It shows the response of neurons tuned to different characteristic frequencies (CFs) through time. We used 512 CFs logarithmically spaced from 250 to 8000 Hz. For each CF, we simulated the average response of 50 AN fibers with different spontaneous rates: high (100 spikes/s), medium (5 spikes/s), and low (0.1 spikes/s), with proportions of 0.6, 0.2, and 0.2, respectively, which correspond to the distribution observed in mammals (Liberman, 1978; Zilany and Bruce, 2007). This neural activity was grouped into time bins of 8 ms and convolved with a 2-sample long rectangular window with a 50% overlap. Figure 3.5 shows example ESNs of reference and target stimuli for different ripple densities.

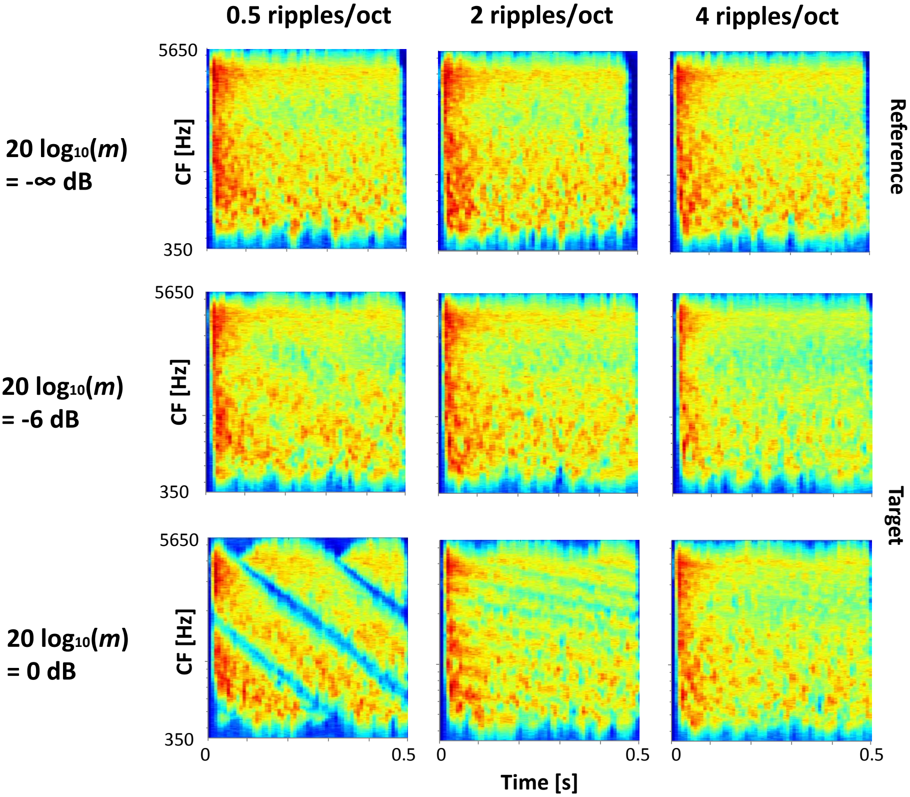

Figure 3.5: ESNs of spectrotemporally modulated stimuli.

3.4.3 Neurogram activity

We quantified the increase of neural activity by computing the mean and standard deviation of the neurograms across different levels. Figure 3.6 shows plots of the ESN neural activity for the reference stimulus ($20\log_{10}(m) = -\infty$ dB) and a fully-modulated target stimulus($20\log_{10}(m) = 0$ dB). In all cases, increasing the level increased the neural activity.

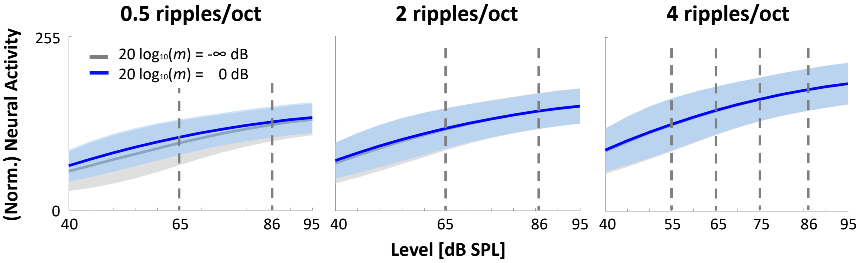

Figure 3.6: Neurogram activity. The solid lines and the shaded areas correspond to the mean and standard deviation, respectively, of each neurogram. This is a measure of the amount of activity at the AN level. A large increase in activity could lead to saturation and, therefore, to a poorer spectrotemporal representation, yielding higher thresholds (Sec. 3.5). The dashed lines represent the levels at which behavioral measurements were obtained.

3.4.4 Neurogram frequency profiling

We defined a *frequency profile* of a neurogram as a slice across its CFs at a given point in time. If we think of a neurogram as an image, a frequency profile would correspond to all the row values of a specific column.

Now, consider the ESNs in Fig. 3.5 for the ripple density of 0.5 ripple/oct. The top ESN ($20\log_{10}(m) = -\infty$ dB) shows a uniform, indistinctive pattern. A frequency profile at any point in time would show a roughly flat curve. In contrast, the bottom ESN ($20\log_{10}(m) = 0$ dB) shows a clear pattern reflecting the spectrotemporal characteristics of the stimulus. A frequency profile at any point in time would show distinct crests and troughs. Therefore, the frequency profiles reflect the spectrotemporal properties of the stimulus coded at the AN level. Figure 3.7 shows frequency profiles for various ripple densities and levels.

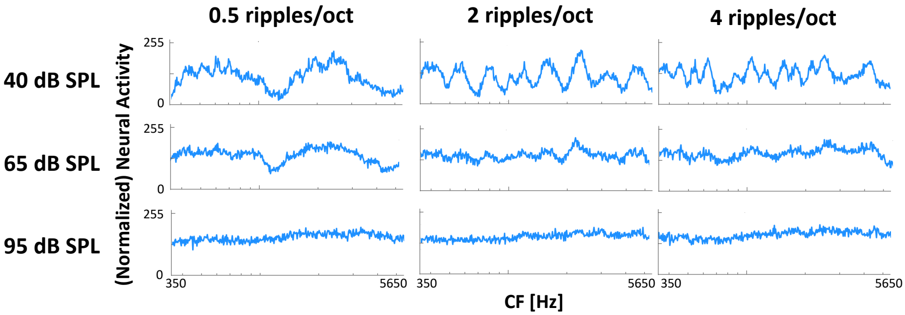

Figure 3.7: Frequency profiles of the ESNs at $t = 250$ ms (total duration of the stimulus was 500 ms).

3.4.5 Dispersion

One measure of the information available at the AN level for detection of modulation is the dispersion of the aforementioned frequency profiles (i.e., columns) of the ESNs across time. The dispersion is a measure of the amplitude of the frequency profile curves. It measures the amount of variation in amplitude across the frequency range. We quantified this dispersion using the interquartile range (IQR), as shown in Eq. 3.3. Furthermore, we also computed a measure of the dispersion variability across all the frequency profiles of a given neurogram, as shown in Eq. 3.4. In both cases, ESNj is the frequency profile at the j-th point in time.

$$\mathrm{ESN_{disp}} = \mathrm{Median}(\mathrm{IQR}(\mathrm{ESN}_j)) \tag{3.3}$$

$$\mathrm{ESN_{dispVar}} = \mathrm{IQR}(\mathrm{IQR}(\mathrm{ESN}_j)) \tag{3.4}$$

Figure 3.8 shows plots of ESN dispersion. Deeper modulations (closer to $20\log_{10}(m) = 0$ dB) led to larger dispersions for lower ripple densities (0.5 and 2 ripples/oct). In all cases, increasing the level reduced the dispersion. This trend was consistent across all three ripple densities.

3.4.6 Regression

Lastly, we related the model results with the behavioral data using a regression model. Since the results of experiment 1 showed that the effect of presentation level was the largest at 4 ripples/oct (and therefore the one that we chose to expand and collect more data), we focused on the behavioral data of experiment 2 (levels of 55, 65, 75, and 86 dB SPL with a ripple density of 4 ripples/oct). Fig. 3.9 shows the ESN representation of the target stimuli with modulation depth corresponding to the behavioral threshold.

We calculated the difference in dispersion between a fully-modulated target stimulus ($20\ \log_{10}(m) = 0$ dB) and the non-modulated reference as a predictor for an exponential regression model as described by Eq. 3.5:

$$\text{Behav. thresh.}(\text{ESN}_{\text{disp}}, \text{ESN}_{\text{disp}}\ \text{ref}) = a\, e^{b\,(\text{ESN}_{\text{disp}} - \text{ESN}_{\text{disp}}\ \text{ref})} \tag{3.5}$$

with parameters a and b. It yielded an (adjusted) R^2 value of 0.98 and a root mean squared error (RMSE) of 0.25 dB. Figure 3.10 shows plots of the behavioral data versus the model metric as well as the regression model. We used the generated model to predict the behavioral thresholds for the different levels. Figure 3.11 shows the model's predictions as well as the mean of the behavioral data (as a reference). The model predictions show that the lowest (best) STMD threshold is around $20\ \log_{10}(m) = -13.5$ dB for the modelled experiment.

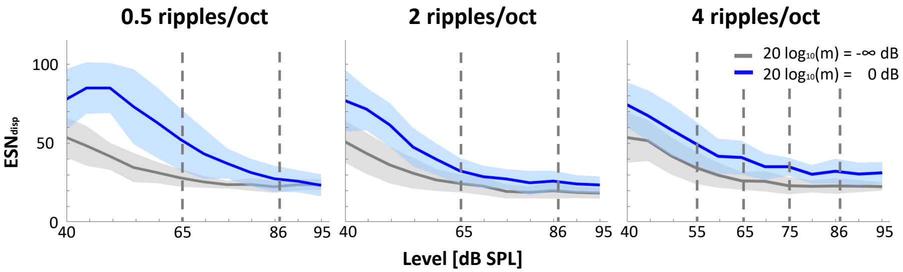

Figure 3.8: Plots of ESN dispersion. The solid lines and the shaded areas correspond to the ESN dispersion and dispersion variability, respectively, for frequency profiles across all time points for each neurogram. The ESN dispersion is a measure of the amount of information for modulation detection available at the AN level (larger dispersion allows for higher detectability, Sec. 3.5). The dashed lines represent the levels at which behavioral measurements were obtained.

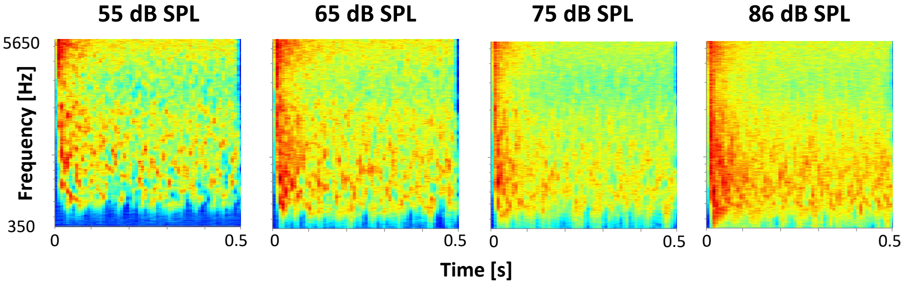

Figure 3.9: ESN representation of stimuli with a modulation depth corresponding to the behavioral threshold. From left to right: -13.2, -12.4, -9.5, and -6.8 dB. In all cases, ripple density was 4 ripples/oct.

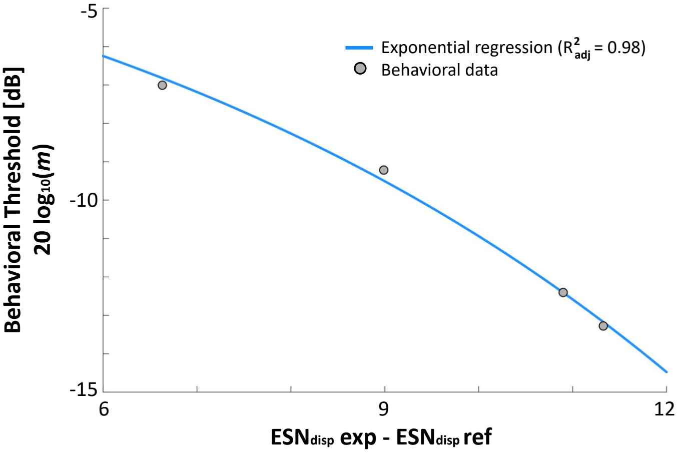

Figure 3.10: Exponential regression model of the behavioral thresholds of experiment 2 (ripple density of 4 ripples/oct).

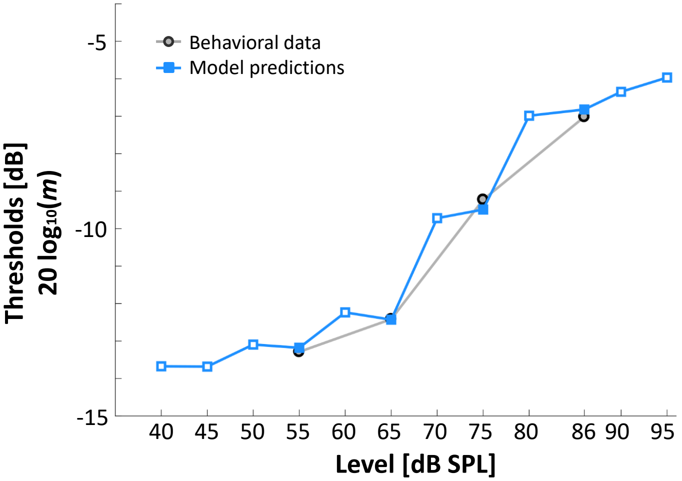

Figure 3.11: Model predictions of the behavioral thresholds across different levels (ripple density of 4 ripples/oct). Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols.

3.5 Discussion

In this study, we investigated how level affected STMD thresholds of young adult NH listeners. We found that higher levels led to increased STMD thresholds. Moreover, increasing ripple density affected the STMD thresholds differently depending on the level. At 65 dB SPL, STMD thresholds were lowest at 2 ripples/oct. In other studies a similar trend was found. Anderson et al. (2012) found lowest thresholds at 3 ripples/oct, followed by increasing thresholds with increasing ripple density (up to 64 ripples/oct). The participants of Eddins and Bero (2007) performed the best either at 2 or 3 ripples/oct. Davies-Venn et al. (2015) found a significant improvement in thresholds from 0.5 to 1 and from 1 to 2 ripples/oct. Other studies (Bernstein and Green, 1987; Leek and Summers, 1996) have also found similar trends. At such level, the most common explanation is that there are two regions at which different cues are being used. For low ripple rates ($<=$ 3 ripples/oct), spectral ripples are being detected using a spectral-contrast mechanism, while for higher ripple rates ($>$ 3 ripples/oct), the spectral cues become weaker and temporal cues become dominant. Our results support this hypothesis. However, further studies are needed to confirm this. At 86 dB SPL, STMD thresholds increased with increasing ripple density, similar to what Bernstein et al. (2013) also found. The effect of presentation level was largest at 4 ripples/oct, where low presentation levels (55 and 65 dB SPL) yielded significantly lower (better) STMD thresholds than high presentation levels (75 and 86 dB SPL).

Understanding the effect of level on STMD thresholds for NH listeners is the first step to understanding it in HI listeners. Although it is very likely that level also affects STMD thresholds of HI listeners, our results cannot be translated directly to the HI population for several reasons. Firstly, increasing the intensity affects the neural saturation of NH and HI listeners differently. This can also affect their perception differently due to their abnormal loudness growth curve (i.e., non-linear loudness shift, Edwards et al., 1998; Hellman, 1999). Additionally, the auditory filters of HI listeners are already abnormally broad (Moore, 2007), resulting in spectral smearing of the stimulus. Furthermore, the large heterogeneity of the HI population (Lopez-Poveda and Johannesen, 2012) would very likely play a role. Therefore, we hypothesize that STMD thresholds of HI listeners will also be affected by level and will be worse than those of NH listeners. However, this would have to be confirmed with further behavioral and modelling studies. This would be a crucial step for further understanding the differences in STMD thresholds between NH and HI participants. Our results show that attributing them only to differences in spectrotemporal sensitivity would be partially true, since level also plays an important role.

We used a computational model with a physiologically-inspired front end to explain the behavioral results (Fig. 3.5). We found that the observed effects of level on the behavioral data could be explained by a worsening of the spectrotemporal representation in the AN because higher levels led to neural saturation “filling in the dips” of the neurograms. This can be seen in the increase of the neural activity (Fig. 3.6) and the flattening of the frequency profiles (Fig. 3.7). Frequency profiles at lower levels reflected the changes of the spectral information across time, while frequency profiles at higher levels lost the representation of this information (Fig. 3.8). All these factors diminish the coding of the spectrotemporal pattern of the modulated stimuli in the AN with increasing level, making it harder to discriminate.

The regression analysis (Fig. 3.10) suggested that information in the auditory periphery is able to account for a large proportion of the variance in the behavioral data, supporting its value for predicting spectrotemporal modulation thresholds in the 4 ripples/oct case (Fig. 3.11). It would be interesting to gather more behavioral data and improve the current model to be able to generalize it to the other conditions (i.e., 0.5 and 2 ripples/oct).

Similar results could have been obtained with a more simple model (e.g., an excitation pattern model, Moore and Glasberg, 1987). However, using frameworks based on the biology of the auditory system presents a few advantages worth mentioning. For instance, since they are representations of the AN responses, they incorporate physiological information inherently. This allows a more direct, transparent understanding of the auditory mechanisms at different stages of the auditory pathway (the periphery in this case), since it gives an insight into the stimulus’s representation at each of these steps. Additionally, the Zilany et al. (2014, 2009) AN model incorporates the effects of sensorineural hearing loss due to damage to the IHCs and OHCs (something that would not be straightforward to do using a non-physiological approach). Now that presented framework has been validated for the NH case, this could be of special interest, since it could allow studying the effect of level on spectrotemporal modulation detection in HI listeners using a similar framework to the one described here.

The effect of presentation level has a number of implications for the use of STMD tests in experimental and clinical environments. When administering STMD tests at different levels, the observed differences in STMD thresholds should (at least partially) be attributed to the effect of level, making it more complex to interpret the contribution of spectrotemporal sensitivity only. In NH participants it is recommended to use a fixed presentation level to allow for direct comparison between their STMD thresholds. However, it is unclear how level affects STMD thresholds in HI listeners. Therefore recommendations for STMD tests in HI participants cannot be made based on our data. Future work will be focused on investigating level effect in different types of spectral and spectrotemporal ripple tests, as well as for HI listeners.

3.6 Conclusions

STMD thresholds were higher (worse) at high than at low presentation levels, with larger differences in thresholds at 4 ripples/oct than at 2 ripples/oct. The computational model with a physiologically inspired front end could account for the behavioral results, showing that information at the peripheral level is sufficient to predict the behavioral thresholds. Therefore, STMD thresholds obtained at different presentation levels are affected not only by differences in spectrotemporal modulation, but also at least partly by level. This needs to be considered when administering STMD tests (both in clinical practice and in experimental research) and when comparing STMD thresholds within and across studies.

CHAPTER 4

Interaural Time Difference Discrimination in Acoustical and Electrical Hearing[1]

---

[1] The work presented in this section has been published as Prokopiou, A., Moncada-Torres, A., Wouters, J. and Francart, T. (**2017**). "Functional modelling of interaural time difference discrimination in acoustical and electrical hearing," Journal of Neural Engineering **14**(4), 1–21. Changes are limited to layout, graphical appearance, and minor editing.

4.1 Abstract

Interaural time differences (ITDs) are important for sound source localisation. In this work, we present a model to predict ITD discrimination thresholds for normal hearing and electric stimulation through bilateral cochlear implants. We combined periphery models of acoustic and electric stimulation with a novel ITD threshold estimation stage, which consists of a shuffled cross correlogram and a binary classifier characterisation method. Furthermore, we present an evaluation framework based on a large behavioral dataset. Our model correctly predicts behavioral observations for unmodulated stimuli (such as pure tones and electric pulse trains) and modulated stimuli for modulation frequencies below 30 Hz. For higher modulation frequencies, the model predicts the observed behavioral trends, but tends to estimate higher ITD sensitivity. The presented model can be used to investigate the implications of modifying the stimulus waveform on ITD sensitivity and to investigate sound encoding strategies.

4.2 Introduction

Cochlear implants (CI) are prosthetic devices that can restore hearing to profoundly deaf persons. Essentially, they consist of a receiver stage and a signal processing stage before electrically stimulating the spiral ganglion using multiple electrodes. Implanting patients with CIs on both ears (i.e., bilateral implantation) is becoming a common clinical practice, particularly in young children. It has been shown that it improves sound localization and speech perception in noise over unilateral CIs (e.g., Firszt et al., 2008; Laback et al., 2015; Offeciers et al., 2005; Van Deun et al., 2009). These benefits arise mostly from utilizing the acoustic head shadow effect, which creates an interaural level difference (ILD) between the two ears (van Hoesel, 2012). Another important cue for sound localization in normal hearing (NH) listeners is the interaural time difference (ITD), which is the most salient cue for sounds with sufficient low-frequency content (Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1992). Furthermore, ITDs have been shown to be dominant for both binaural unmasking and attention-driven spatial release from masking (Bronkhorst, 2000; Bronkhorst and Plomp, 1988; Kidd et al., 2010).

Unfortunately, ITD cues are not fully coded by clinical CI sound processors. Furthermore, bilateral cochlear implant (BiCI) users have been shown to have poor sensitivity to high-rate pulse train ITDs (Laback et al., 2015; Noel and Eddington, 2013). However, behavioral studies have given evidence to support the idea that modifying the modulating envelope of a carrier tone has the potential to influence ITD perception in BiCI users (Laback et al., 2011; Noel and Eddington, 2013; van Hoesel et al., 2009) and in bimodal users (i.e., one ear with a hearing aid and the other with a CI, Francart, Brokx and Wouters, 2009; Francart et al., 2011, 2014). There are still unknowns about ITD perception when using a CI. Specifically, how do temporal properties of the stimulation envelope (such as modulation frequency and depth), temporal gaps, and the rate of change of envelope amplitude affect ITD perception.

Various binaural models have been developed which attempt to address these questions by describing the complex interactions between perceived location due to ITD and stimulus waveform. These models can be classified into two categories. On one hand, *biophysical models* aim to describe particular neural pathways by characterising underlying binaural mechanisms (Chung et al., 2014; Rothman and Manis, 2003; Wang and Colburn, 2012; Wang et al., 2014). On the other hand, *statistical models* aim to describe a generic binaural processor using signal processing techniques that relate empirical observations to the model outcomes (Bernstein and Trahiotis, 2002, 2009; Breebaart et al., 2001; Colburn, 1973, 1977; Dietz et al., 2011; Pulkki and Hirvonen, 2009; Takanen et al., 2014). However, this type of models typically do not attempt to explain the underlying mechanisms behind said observations.

Furthermore, a substantial body of work already exists for describing the acoustic and the electric stimulation of the auditory nerve (AN). In the acoustic case, the model proposed by Zilany et al. (2009) for acoustical stimulation can accurately predict various temporal phenomena such as non-linear tuning, level-dependent phase, compression, suppression, shift in the best frequency as a function of level, adaptation, as well as some other non-linearities seen at high sound levels based on several AN datasets. Additionally, its most recent version corrected the saturation of firing rates of higher characteristic frequency fibres when stimulated by low frequency tones (Zilany et al., 2014). This modelling effort improved the prediction of AN responses to a wide variety of complex sounds (such as amplitude-modulated stimuli) and forward-masking paradigms (Zilany et al., 2009), while accounting for long-term dynamics of AN responses (Zilany and Carney, 2010).

In the electric case, the situation is rather different. There is a broader landscape of models which can be separated into three main categories (Nicoletti et al., 2013). *Point neuron models* (Bruce et al., 1999; Goldwyn et al., 2012; Mino et al., 2002; Motz and Rattay, 1986; Rattay, 1986) aim at modelling individual neuron detailed dynamic properties. *Multi-compartmental models* (Briaire and Frijns, 2000; McNeal, 1976; Mino et al., 2002; Woo et al., 2010) work as extensions of point neuron models by considering how and where the action potentials are generated on the AN following electrical stimulation. They are typically used for connecting a sequence of neurons. Finally, *Population models* (Nicoletti et al., 2010) aim at replicating excitation patterns along the entire cochlea.

The complex relation between stimulus waveform and ITD perception is not yet fully understood. In order to describe human ITD perception, we developed a computational model relying on the working hypothesis that central auditory processing is normal (i.e., not impaired) for BiCI users, which is also adopted by Chung et al. (2014). As such, a single model is used to process the AN response for both NH and BiCI users by utilizing their respective acoustic and electric stimuli at AN level as inputs. For the acoustic case we chose the aforementioned model proposed by Zilany et al. (2014) to represent the neural responses of the acoustic stimulation peripheral pathway. For the electric case, we considered direct stimulation of the spiral ganglion without any residual hearing. Furthermore, we considered a single bilateral electrode pair stimulation in order to reduce the number of assumptions that would be necessary to account for across channel integration phenomena. Thus, a point neuron approach is sufficient, since capturing temporal details of the dynamic response of the AN is vital to single electrode pair stimulation. In particular, we chose the model proposed by Goldwyn et al. (2012), which utilises point-process analysis to modify a parameter space to tune neuron firing parameters, such as response latency, threshold, relative spread, jitter, and summation time. The model includes temporal filtering that represents sub-threshold dynamics of the membrane potential, a non-linearity associated with spike generating processes, and a secondary filter that accounts for variability in spike timing. It incorporates dynamical and stochastic properties that are important to high pulse rate stimulation, which is particularly relevant to the clinical stimuli used in modern CI devices that can reach stimulation rates in the range of 1,000 - 20,000 pulses per second (PPS). Both the acoustic and electric front end were combined with a novel neurometric psychometric estimation method. We validated the model by comparing its ITD thresholds (i.e., just noticeable difference, JND) predictions with relevant psychoacoustic experimental results from the literature.

4.3 Materials & Methods

The framework of the proposed computational model is illustrated in Fig. 4.1. A particular pair of left and right ear stimuli were given as input to a model of the human auditory periphery. The resulting output of the periphery described the neuronal activity of the AN for either acoustic or electric stimulation. This neuronal activity was quantified as a temporal pattern of action potentials and was used as an input to the next stage of the model: the shuffled cross correlogram (SCC, Joris et al., 2006). This stage gave an estimate of the relative timing disparity between the left and the right AN activity.

The left and right ear stimuli used as inputs to the model were either presented with an ITD equal to 0, serving as a reference, or with an ITD that is non-zero, serving as a target. Both the target and the reference were used as input to a binary classifier characterization stage where the ITD threshold was estimated.

The model parameters are the neural density distribution on the basilar membrane, the bin-width of the SCC, and jitter introduced after the peripheral neuronal activity estimation. The remaining parameters particular to the peripheral models were adopted from their respective publications (Goldwyn et al., 2012; Zilany et al., 2014). Specifically, for the acoustic stimulation the parameters of the Zilany et al. (2014) model are the functioning of the outer and inner hair cells, which was set to normal (i.e., healthy function); the species model, which was set to be human with basilar membrane tuning from Shera et al. (2002); the AN spontaneous rate which was set to be a high spontaneous rate; the noise type which was set to be variable (i.e., different every simulation); and the implementation of the power-law function which set to be the actual implementation (i.e., not an approximation). Furthermore, the binsize for the resulting PSTH was set to be equal to the sampling time to get the individual timing for each action potential. The individual time is important for the calculation of the SCC (Sec. 4.3.3). For the electric stimulation the parameters of the Goldwyn et al. (2012) model were set to a threshold of 0.852 mA (Miller et al., 2001), a relative spread of 4.87% (Miller et al., 2001), a jitter of 85.5 μs (Miller et al., 2001), and a summation time of 250 μs (Cartee et al., 2006).

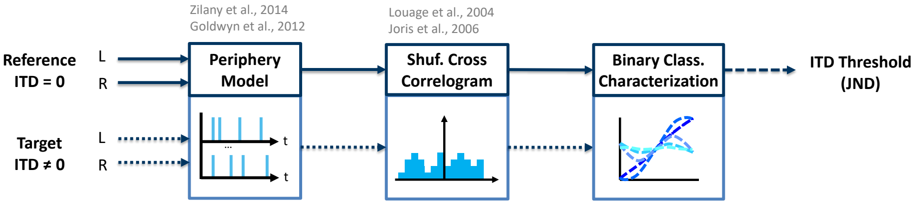

Figure 4.1: Block diagram of the proposed model framework. Solid arrows represent the reference case (ITD = 0 μs). Dotted arrows represent the target case (ITD $\neq$ 0). The dashed-and-dotted arrow represents the ITD threshold (i.e., just noticeable difference, JND).

4.3.1 Stimuli

The proposed computational model aims to estimate ITD sensitivity for both acoustic and CI stimulation. All the stimuli were previously described in various behavioral studies where ITD thresholds were reported. The selection of these studies aims to create an evaluation framework for the model.

Acoustic stimuli

For the NH situation, the subject was assumed to have normal acoustic hearing on both ears. The setup of the model allowed estimation of ITD thresholds narrow band stimuli. The peripheral model operated within a bandwidth of one equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990). Corresponding to the behavioral data that was used, the ITD was static and there was no ILD. Furthermore, the behavioral studies considered ongoing cues, so the onset and offset cues were reduced using a 20 ms cosine square ramp for all acoustic stimuli.

The model's predictions were compared against behavioral data for pure tones (Brughera et al., 2013), sinusoidally amplitude modulated (SAM) tones (Bernstein and Trahiotis, 2002), transposed tones (Bernstein and Trahiotis, 2002), raised sine tones both as the modulation frequency and the modulation depth is varied (Bernstein and Trahiotis, 2009), and finally as the tone is modulated with trapezoidal shaped waveforms (Laback et al., 2011). All the acoustic stimuli parameters are shown in Table 4.1.

Table 4.1: NH stimuli parameters. All stimuli were generated with a sampling frequency of 100 kHz. f = frequency, f_c = carrier frequency, f_m = modulation frequency, m = modulation depth, n = exponent, SAM = sinusoidally amplitude modulated.

Type of tone	Parameters	Level [dB SPL]	Duration [ms]
Pure	$f = \{250, 500, 700, 800, 900, 1000, 1200, 1250, 1300\}$ Hz	75	300
SAM	$f_c = \{4000, 6000, 10000\}$ Hz $f_m = \{30, 60, 120, 250, 500\}$ Hz $m = 1$	75	300
Transposed	$f_c = \{4000, 6000, 10000\}$ Hz $f_m = \{30, 60, 120, 250, 500\}$ Hz $m = 1$	75	300
Raised sine (f_m)	$f_c = 4000$ Hz $f_m = \{30, 60, 120, 250\}$ Hz $m = 1$ $n = 1, 2, 4, 8$	75	300
Raised sine (m)	$f_c = 4000$ Hz $f_m = 128$ Hz $m = \{0.25, 0.5, 0.75, 1\}$ $n = 1, 2, 8$	75	300
Trapezoid	$f_c = 8727$ Hz $f_m = 27.3$ Hz $m = 1$ slope = $\{6, 8, 10, 12\}$ dB/ms off time = $\{1, 6, 12, 18, 21\}$ ms	78.2	1000

Electrical Stimuli

For the BiCI situation, the subject was assumed to have no residual hearing and a CI implanted in both ears. The electric pulses for all the electric stimuli used were biphasic with a phase duration of 25 μs and an interphase gap of 8 μs. The pulse rate was variable, ranging from 40 to 5000 PPS (depending on the behavioral data). Defining the stimulation current amplitude is not a straightforward task as the connection between loudness perception and stimulating current is highly subject dependent. In order to quantify the CI stimulation intensity, we used the firing efficiency measure, which is explained as follows.

The firing efficiency curve is an input/output function that relates the current level of a single pulse of current to the probability that the stimulus evokes a spike. The definition of threshold current (I_{thr}) for a particular neuron is when the probability of generating an action potential is equal to 50%. The simulated neuron had a $I_{thr} = 0.852$ mA. The peak current (I_{peak}) for each experimental condition was set to have a firing efficiency of 1.25 dB with reference to I_{thr}, which means that $I_{peak} = I_{thr} * 10^{\frac{1.25-1}{20}} = 0.877$ mA. Using the firing efficiency curve, this I_{peak} value corresponded to ~75% chance of generating an action potential from a single pulse. The stimulation current in all the electric stimuli used was scaled to have the maximum current be equal to the calculated peak current.

The firing efficiency of 1.25 dB was chosen because the average firing rate of neurons activated with electric stimulation was found to be comparable to acoustic stimulation at 75 dB SPL. We compared them by estimating the baseline firing rate of either, and varying the firing efficiency until the baseline firing rate matched. The baseline firing rate was estimated as a running average of spikes per neuron over a 15 ms time window for low rate stimulation of constant envelope (i.e., unmodulated) stimuli.

The binaural model was compared against behavioral data for ITD sensitivity of unmodulated low-frequency pulse trains (Egger et al., 2016; Laback et al., 2007; van Hoesel, 2007; van Hoesel and Clark, 1997; van Hoesel et al., 2009; van Hoesel and Tyler, 2003), SAM pulse trains (Noel and Eddington, 2013), and trapezoidally modulated pulse trains (Laback et al., 2011). All the electrical stimuli parameters are shown in Table 4.2.

Table 4.2: BiCI stimuli parameters. All stimuli were generated with a sampling frequency of 100 kHz. SAM = sinusoidally amplitude modulated, PPS = pulses per second, f_m = modulation frequency, m = modulation depth.

Type of pulse train	Parameters	Firing efficiency [dB]	Duration [ms]
Unmodulated	Pulse rate= $\{40, 100, 200, 300, 400$ $500, 600, 700, 800, 900, 1000\}$ PPS	75	300
SAM	Pulse rate= 1000 PPS $f_m = \{4, 8, 16\}$ Hz Pulse rate= 5000 PPS $f_m = \{50, 100, 200, 500\}$ Hz $m = 1$	75	300
Trapezoidally modulated	Pulse rate= 1515 PPS $f_m = 27.3$ Hz $m = 1$ slope = $\{6, 8, 12\}$ dB/ms off time = $\{1, 6, 12, 18\}$ ms	78.2	1000

4.3.2 Peripheral Model

The peripheral model stage emulated the physiological conversion of an external stimulus (either acoustic or electric) to a binary sequence of action potentials on the AN. We used two identical peripheral models for the left and the right ears.

The number of acoustical neurons simulated N_{acou} was set equal to the amount of neurons that spanned one ERB (as described by Glasberg and Moore, 1990) and defined in Eq. (4.1). The centre frequency (CF) of the band was chosen to be equal to the carrier frequency of the stimulating tone.

$$\mathrm{ERB} = 24.7\left(\frac{4.37\mathrm{CF}}{1000} + 1\right) \tag{4.1}$$

The calculation of N_{acou} is shown in Eq. (4.2), where ρ is the density of neural innervation of the human basilar membrane as a function of the distance along the basilar membrane x as described by Spoendlin and Schrott (1988). Specifically, $\rho(x)$ was calculated using a polynomial fit to their data to interpolate the measurements required by the model. x_{low} and x_{high} are the corresponding locations on the basilar membrane of humans as described by Greenwood (1990) and defined in Eq. 4.3 and 4.4, respectively.

$$N_{acou} = \int_{x_{low}}^{x_{high}} \rho(x)\mathrm{d}x \tag{4.2}$$

$$x_{low} = \frac{1}{2.1}\log_{10}\left(\frac{CF - \frac{\mathrm{ERB}}{2}}{165.4} + 0.88\right) \tag{4.3}$$

$$x_{high} = \frac{1}{2.1}\log_{10}\left(\frac{CF + \frac{\mathrm{ERB}}{2}}{165.4} + 0.88\right) \tag{4.4}$$

The equivalent consideration for the electric case is the spread of excitation. However, it was not considered in this work because only one electrode was simulated in each cochlea. The number of electrically stimulated neurons (N_{elec}) was fixed at 1000, which was similar on average to N_{acou} (i.e., the amount of neurons that roughly span one ERB).

The output of the action potential estimating models (both acoustic and electric) was modified by introducing temporal jitter with a standard deviation of 250 μs. This was done to simulate noise addition from action potential propagation (Faisal et al., 2008) and to remove the phase locking to the pulsatile electric stimulus to emphasise the envelope.

4.3.3 Shuffled Cross Correlogram

The Shuffled Cross Correlogram (SCC, Joris et al., 2006) is a variation of the Shuffled Auto Correlogram (SAC) proposed by Joris (2003). The SCC functions as a binaural coincidence detector by comparing the firing timing between neurons of the left and right auditory fibers (Fig. 4.2). It was computed as follows. First, N spike trains from the left ear and N spike trains from the right ear were fed as inputs. Then, the forward and backward time intervals between all the spikes of the first left spike train and all the spikes of all the right spike trains were measured. The same was done with all the spikes of the second left spike train and all the spikes of all the right spike trains and so on. These time intervals were tallied into a histogram. For the latter, a standard binwidth size $\Delta\tau$ of $50\,\mu$s was used, as suggested by Louage et al. (2004). A schematic representation of the SCC computation is shown in Fig. 4.2. The operation of counting intervals can be thought as an equivalent of counting coincident spikes between two different spike trains. Therefore, the SCC can be plotted as the number of intervals (or counts) versus the delay value (Louage et al., 2004).

Since we were interested in the firing temporal properties, the SCC was normalized by N, average firing rate r, $\Delta\tau$, and stimulus duration D. This was done by dividing the SCC by the term $N^2\, r^2\, \Delta\tau\, D$, making it dimensionless and independent from these parameters. Thus, a count value larger than 1 means that spikes tend to be correlated between spike trains; a count value of 1 shows a lack of stimulus-induced temporal correlation; a count value smaller than 1 indicates anticorrelation (Joris et al., 2006).

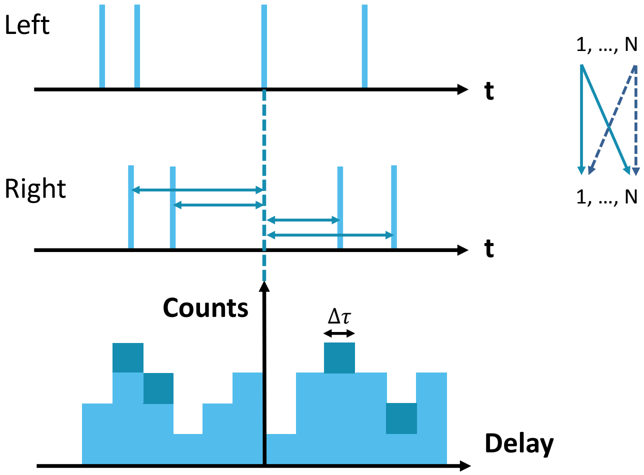

Figure 4.2: Schematic representation of the Shuffled Cross Correlogram (SCC) computation. Notice how the SCC can be plotted as the number of intervals (or counts) versus the delay value.

4.3.4 Binary Classifier Characterisation

The SCC yielded a distribution which characterises the joint temporal properties of the action potentials generated at both ANs. The SCC distribution translates along the x-axis as the ITD between left and right stimuli varies (Fig. 4.3, panel A). Therefore, the purpose of the binary classifier characterisation (BCC) method was to produce a metric which quantified the mismatch between a reference distribution (ITD $=$ 0, i.e., non-shifted) and a target distribution (ITD $\neq$ 0, i.e., shifted) to provide an estimation of the ITD threshold.

The mismatch between the reference and the target distributions was quantified in an analogous way to the computation of the area under the curve (AUC) in a receiver operating characteristic (ROC) curve. There are some differences, which are explained as follows. Typically, when computing the ROC curve, the ordinate is the true positive rate and the abscissa is the false positive rate. In the BCC method, the cumulative surface area of the reference distribution ($\mathrm{CS}_R(\sigma)$), is the ordinate, and the cumulative surface area of the target distribution ($\mathrm{CS}_T(\sigma, \mathrm{ITD})$) is the abscissa; where σ is the SCC lags. The cumulative surface areas of the two distributions are computed within an integration window w as shown in Fig. 4.3 (panel A).

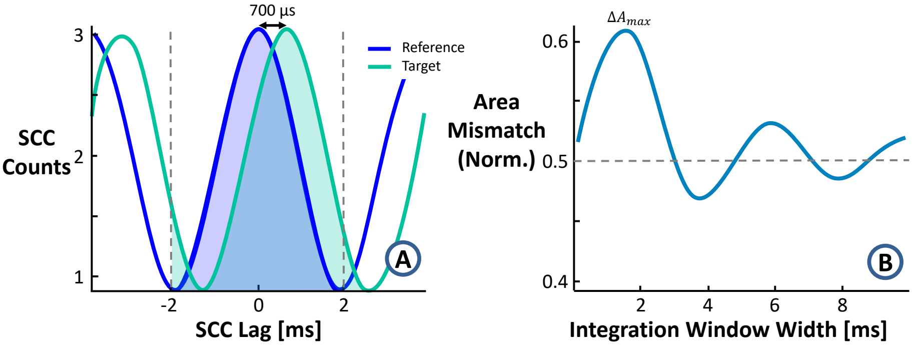

Figure 4.3: Panel A. SCC curve for a pure tone with frequency of 250 Hz at 75 dB SPL with a reference ITD = 0 and target ITD = 700 μs. Vertical dashed lines delimit the integration window w. The BCC method relates the difference of the shaded areas.
Panel B. Plot of $\Delta\mathrm{A}(w)$ when ITD = 700 μs, as defined in Eq. (4.5). It is a measure of the normalised area mismatch between the reference and target SCC of panel A. The dashed line at 0.5 indicates the point where the two areas are equal.

The AUC was computed for various integration window widths. It is defined in Eq. (4.5) as $\Delta\mathrm{A}(w, \mathrm{ITD})$ where w corresponds to the width of the integration window of coincidence detection neurons in the medial superior olive (MSO).

$$\Delta\mathrm{A}(w, \mathrm{ITD}) = \int_{-w}^{w} \mathrm{CS}_R(\sigma) \frac{\partial \mathrm{CS}_T(\sigma, \mathrm{ITD})}{\partial \sigma} \mathrm{d}\sigma \tag{4.5}$$

An exploration of $\Delta\mathrm{A}(w)$ as w changes showed that it resembled a dampened oscillation (Fig. 4.3, panel B), where its maximum value $\Delta\mathrm{A}_{max}$, defined $\psi(\mathrm{ITD})$ as shown in Eq. (4.6).

$$\psi(\mathrm{ITD}) = \Delta\mathrm{A}(w_{max}, \mathrm{ITD}) \tag{4.6}$$

Selecting the maximum value of $\Delta\mathrm{A}(w, \mathrm{ITD})$ as w eliminated the variable w by fixing it as the constant w_{max} shown in Eq. (4.7).

$$w_{max} = \mathrm{argmax}_w(|\Delta\mathrm{A}(w, \mathrm{ITD})|) \tag{4.7}$$

Therefore, $\psi(\mathrm{ITD})$ varies only as ITD changes for a particular stimulus. For example, Fig. 4.3 (panel B) shows that $\psi \approx 0.61$ (labelled with $\Delta\mathrm{A}_{max}$) for a pure tone with frequency of 250 Hz and an ITD = 700 μs. Fig. 4.4 (panel A) shows examples with different ITDs. $\psi(\mathrm{ITD})$ increases with increasing ITD.

The $\psi(\mathrm{ITD})$ metric can be thought of as a neurometric-psychometric method which applies signal detection theory to map neural activity to stimulus sensation (Stüttgen et al., 2011). The dependence of $\psi(\mathrm{ITD})$ on ITD is illustrated in Fig. 4.4 (panel B), where we observe that $\psi(\mathrm{ITD})$ is a neurometric function derived from the BCC analysis. The $\psi(\mathrm{ITD})$ has certain properties which make it comparable to a psychometric function for the region of physiologically relevant ITD values. For instance, when the reference and target distributions overlap, $\psi(\mathrm{ITD}) = 0.5$. It is normalised such that $0 < \psi(\mathrm{ITD}) < 1$. Lastly, there is a logistic function whose inflection point lies on the threshold (i.e., when the reference and target distributions overlap at ITD = 0).

The next step was to relate $\psi(\mathrm{ITD})$ with an estimate of the threshold (i.e., JND) as shown in Eq. (4.8). Since $\psi(\mathrm{ITD})$ resembles a psychometric function, the threshold was defined as the inverse of the slope at ITD = 0 (insert of Fig. 4.4, panel B). A common threshold measure is defined by the non-zero value of a varied parameter (here the ITD) where the psychometric function reaches a certain (defined) performance level. The steeper the slope at the threshold (i.e., the inflection point of the logistic function) the less change in the varied parameter is needed (i.e., lower threshold) before the defined performance level is attained. As such, the slope is inversely proportional to the threshold value. This relation is described in Eq. (4.8). It produced a dimensionless measure of ITD threshold and thus receives the arbitrary quantity of *model units*.

$$\mathrm{JND}(\mathrm{ITD}) = \left(\frac{\mathrm{d}\psi(\mathrm{ITD})}{\mathrm{dITD}}\right)^{-1} \tag{4.8}$$

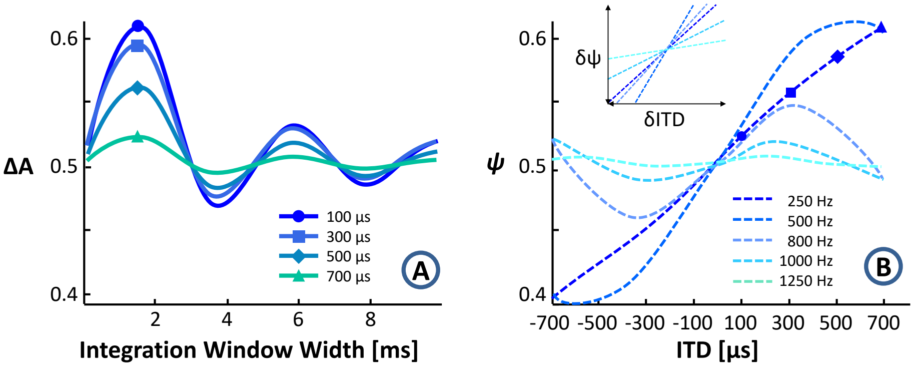

Figure 4.4: Panel A. Various $\Delta\mathrm{A}(w, \mathrm{ITD})$ curves exemplifying the AUC computation (Eq. (4.5)) as a function of the integration window width (w) and various ITD values. The stimulus was a pure tone with frequency of 250 Hz at 75 dB SPL. The symbols ($\circ, \square, \diamond, \triangle$) show the $\Delta\mathrm{A}_{max}$ of the particular curve and correspond to the ψ(ITD) value on panel B.
Panel B. Neurometric-psychometric ψ(ITD) curves of pure tone stimuli of various frequencies as a function of ITD. The insert shows the slope estimation at ITD $= 0$, where the slope is approximated as $\delta\psi/\delta\mathrm{ITD}$.

Note that for the threshold estimations presented in Sec. 4.4, the reference ITD $= 0$ (Bernstein and Trahiotis, 2002, 2009; Brughera et al., 2013; Laback et al., 2011). Thus we define the threshold as the minimum detectable change for said reference.

4.3.5 Model Performance Evaluation

Pearson correlation coefficients were computed between the behavioral and the model predictions to compare them objectively for each dataset. Linear scale was used to order the behavioral data and model predictions in μs and model units, respectively.

4.4 Results

4.4.1 Acoustical Hearing

Pure Tones

Figure 4.5 shows a direct comparison of the model predictions and behavioral data. Note that the scales are measures of different quantities, specifically for human behavioral responses the ITD threshold is given in μs, whereas the model has arbitrary model units. We observe that the high frequency limit for ITD detection reported from Brughera et al. (2013) is consistent with the model ITD threshold predictions. However, we also observe that the low frequency ITD detection threshold is estimated by the model to be lower than what the behavioral data indicates. The correlation between the model prediction and the behavioral data of 0.91 ($p < 0.01$) indicates a good correspondence between the model predictions and behavioral data.

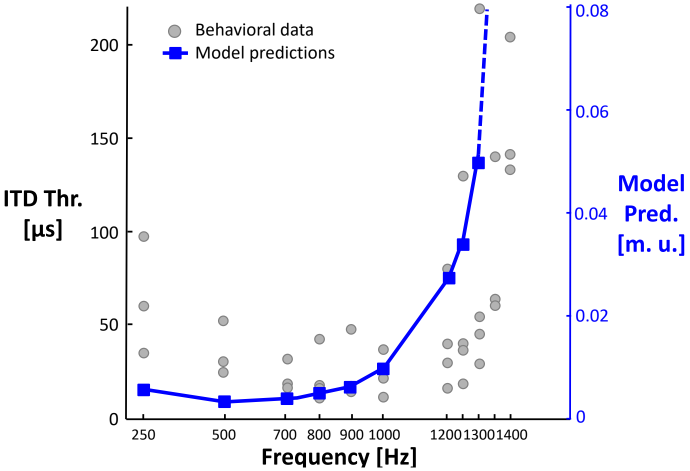

Figure 4.5: Pure tone model performance. Experimental data from Brughera et al. (2013). The stimulus duration was 300 ms at an intensity of 75 dB SPL. Note that the term *model units* used here and throughout the manuscript represents a dimensionless measure of ITD threshold, which is consistent across behavioral data for various stimuli. The threshold estimation is inaccurate above 1300 Hz (where the high frequency limit exists), which is shown in dotted lines.

Sinusoidally Amplitude Modulated Tones

Figure 4.6 shows a direct comparison between behavioral and model data. The model correctly predicts the trend of human behavioral responses for variation in the modulation frequency of SAM tones (Bernstein and Trahiotis, 2002). Specifically, ITD thresholds go up for low and high f_m and they reach a minimum in the region of 100—200 Hz. Note that for all f_c the high frequency cut-off is predicted to be not as steep as behavioral data indicate. Moreover, specifically for $f_c = 4$ kHz, there is some extent of f_m mismatch between the minimum ITD threshold for human behavioral responses and the minimum ITD threshold estimated by the model. Additionally to visual inspection, this frequency mismatch was identified with a change in the correlation value. If we maintain equivalence between the modulation frequency axis between model and data we get a correlation value -0.283. However when we translate the model outcome along the f_m axis to align the minima of model and data we get a correlation value of 0.89 ($p = 0.11$). For $f_c = 6$ kHz, we do not observe frequency mismatch, which is indicated by a strong correlation of 0.89 ($p = 0.04$). For $f_c = 10$ kHz, we observe a good match for low modulation frequencies ($\leq$ 125 Hz), whereas for higher modulation frequencies the model predicts higher sensitivity (i.e., lower ITD threshold values). We also observe that when changing f_c the model does not predict the data (Bernstein and Trahiotis, 2002). Behaviorally, an increase of f_c increases ITD thresholds. The model is unable to capture this change. This shortcoming is also reflected in a lower correlation value (0.58, $p = 0.04$) between behavioral data and model predictions across all f_c.

Transposed Tones

Figure 4.7 shows behavioral and model data. Comparably to SAM tones, we observe again a similar trend for both the model and the behavioral data where the ITD threshold is minimised for a particular f_m region. This f_m region is predicted to be lower than the behavioral data (Bernstein and Trahiotis, 2002). The overall trend for individual carrier frequencies is well predicted with correlation values of $r = 0.81$ ($p = 0.09$), $r = 0.87$ ($p = 0.05$) and $r = 0.93$ ($p = 0.02$) for centre frequencies of 4 kHz, 6 kHz and 10 kHz respectively. However, similarly to the SAM tone case, the across centre frequency variation is not well predicted with a correlation of $r = 0.55$ ($p = 0.07$).

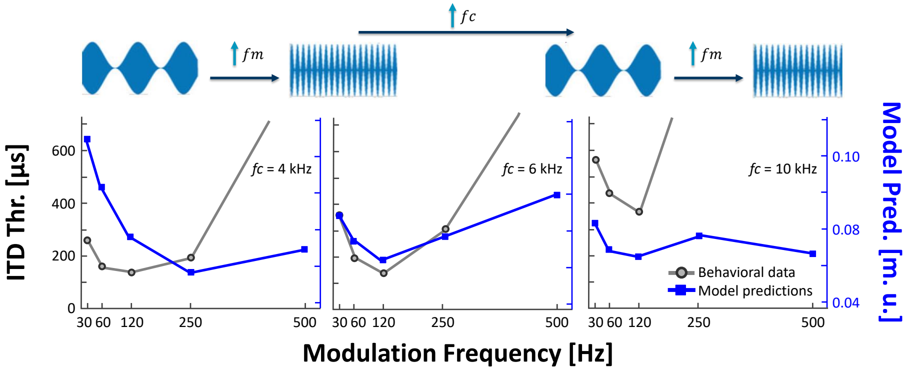

Figure 4.6: SAM tones model performance. Experimental data from Bernstein and Trahiotis (2002). The stimulus duration was 300 ms at an intensity of 75 dB SPL. The change in f_c essentially translates the region of neuronal stimulation on the basilar membrane.

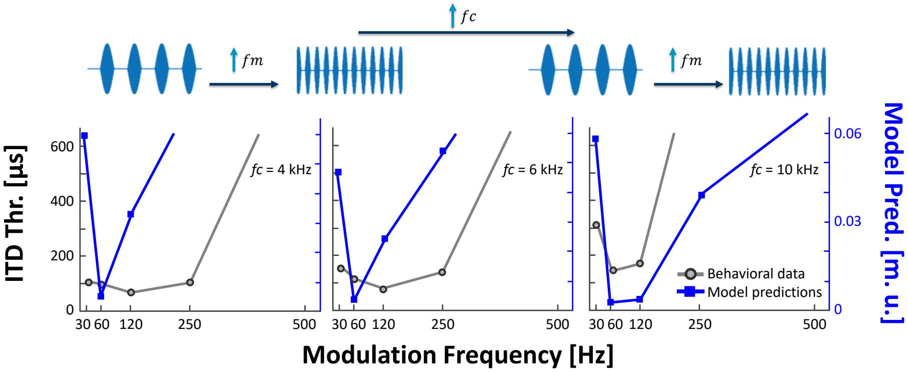

Figure 4.7: Transposed tones model performance. Experimental data from Bernstein and Trahiotis (2002). The stimulus duration was 300 ms at an intensity of 75 dB SPL.

The modelling outcomes indicate that the ITD threshold is lower for transposed stimuli compared with SAM tones, which agrees with reported observations (Bernstein and Trahiotis, 2002). However, the model predicts a larger relative difference between transposed tones and SAM tones (Fig. 4.8).

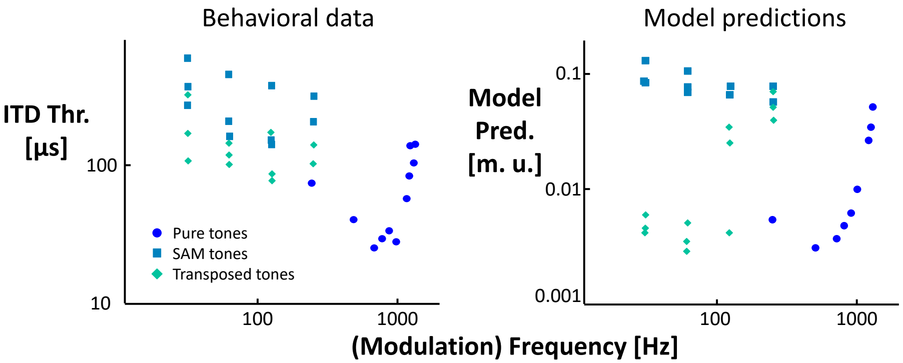

Figure 4.8: Scatter plot of the mean of behavioral data (left panel) and model prediction (right panel) for pure tones (Brughera et al., 2013), SAM tones (Bernstein and Trahiotis, 2002), and transposed tones (Bernstein and Trahiotis, 2002).

Raised sine modulation

Figure 4.9 shows a direct comparison of model prediction and behavioral data. Behavioral measures indicate a decrease in ITD thresholds as the modulation exponent is increased for various modulation frequencies (Bernstein and Trahiotis, 2009), which is predicted by the model with a correlation of $r = 0.60$ ($p = 0.01$), across exponent values. Similarly to the previous modulated tones the behavioral data indicate a certain modulation frequency region where the ITD threshold is minimized. However, similarly to the transposed tones, the frequency region which minimises ITD thresholds is predicted by the model to be lower than the behavioral data reported (Bernstein and Trahiotis, 2009). If we account for this frequency mismatch by shifting the model prediction along the f_m axis then the correlation values for the individual exponent values are $r = 0.88$ ($p = 0.12$), $r = 0.98$ ($p = 0.02$), $r = 0.77$ ($p = 0.22$) and $r = 0.98$ ($p < 0.01$) for exponent equal to 1, 2, 4 and 8 respectively, indicating a good overall trend prediction.

Additionally, the effect of decreasing the modulation depth was investigated, as shown in Fig. 4.10. The behavioral data show a reduction in ITD threshold as the modulation depth increases, which is also indicated by the model, with correlation values of $r = 0.89$ ($p = 0.11$), $r = 0.94$ ($p = 0.06$), and $r = 0.57$ ($p = 0.43$) for exponents 1, 2, and 8 respectively. However as the exponent increases for the various modulation depths, the model predicts that the ITD threshold increases, which is contrary to the reported data (Bernstein and Trahiotis, 2009), and is reflected in a low correlation value of $r = 0.46$ ($p = 0.13$).

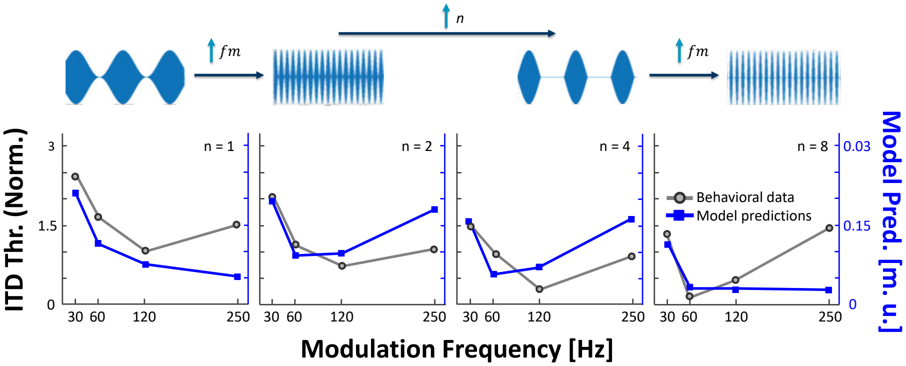

Figure 4.9: Model performance for raised sine with variable modulation frequency. Behavioral data from Bernstein and Trahiotis (2009). The stimulus duration was 300 ms at an intensity of 75 dB SPL. The normalization in the behavioral data was accomplished by dividing an individual listener's threshold by the same listener's threshold at 128 Hz modulated SAM tone. This was done so as to remove inter-subject variability (Bernstein and Trahiotis, 2009).

Trapezoidal Modulation

Figure 4.11 illustrates a direct comparison between model predictions and behavioral data. The behavioral data reported indicate that by increasing the off time, the ITD detection thresholds decrease (Laback et al., 2011). The model output correctly indicates the same trend as observed in the behavioral data, specifically the correlation value between model prediction and behavioral data is $r = 0.97$ ($p = 0.03$), $r = 0.95$ ($p = 0.01$), $r = 0.93$ ($p = 0.02$), and $r = 0.94$ ($p = 0.02$) when the slope is 6, 8, 10 and 12 dBms^{-1} respectively. A further observation was that by increasing the rising and falling slope the ITD detection thresholds decrease (Laback et al., 2011). When comparing across all the slope conditions the correlation value is $r = 0.82$ ($p < 0.01$), which indicates good prediction from the model, albeit a higher sensitivity to slope changes is indicated by the model than the behavioral data.

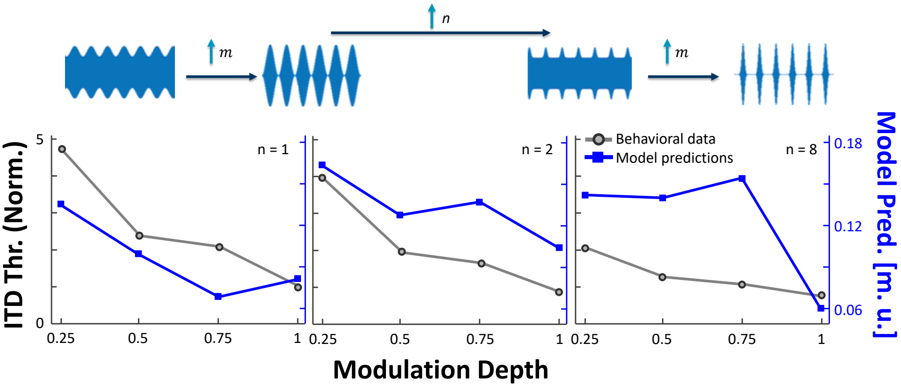

Figure 4.10: Model performance for raised sine with variable modulation depth. Experimental data from Bernstein and Trahiotis (2009). The stimulus duration was 300 ms at an intensity of 75 dB SPL. The normalization of the behavioral data is done as described in the caption of Fig. 4.9.

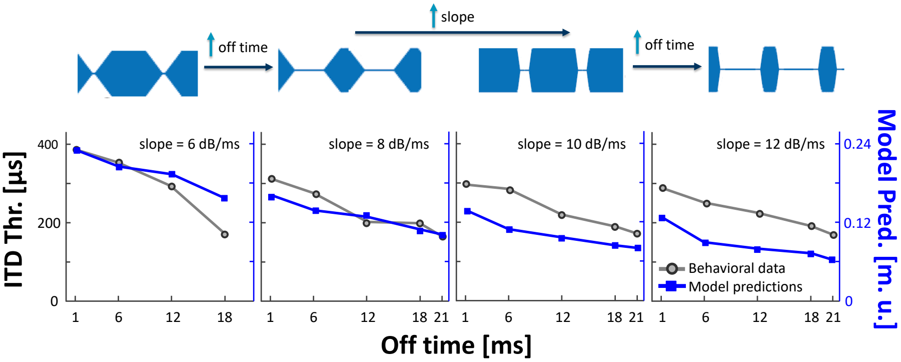

Figure 4.11: Model performance for trapezoid modulation. Experimental data from Laback et al. (2011). The stimulus duration was 1 s at an intensity of 75 dB SPL.

4.4.2 Electrical Hearing

Unmodulated low frequency pulse trains

ITD perception was tested as a function of stimulation rate (in PPS). Figure 4.12 shows a comparison of model prediction against behavioral data.

The behavioral data indicate that for rates higher than 100 PPS, the majority of BiCI users start having larger ITD detection thresholds. which become unmeasurable for rates higher than 400—800 PPS (Laback et al., 2015). The model correctly predicts that for a certain region of low rate stimulation the ITD detection threshold is constant and it starts increasing after a certain point and becomes physiologically non-relevant for rates higher than 800 PPS. However, it estimates lower ITD detection thresholds than the behavioral data in the region of 400—800 PPS. Overall, the model predicts well the behavioral trend with a correlation value of $r = 0.91$ ($p = 0.002$) between model predictions and behavioral data.

Sinusoidally amplitude modulated (SAM) pulse trains

Figure 4.13 shows the comparison of behavioral data with the model prediction. The behavioral result reported was a v-shaped curve of ITD threshold versus f_m (Noel and Eddington, 2013). This is somewhat predicted by the model with the presence of the notch, however the notch has a different shape and there is a mismatch of the frequency at the minimum point. The behavioral data indicate an ITD threshold minimum at around 100 Hz, while the model predicts such minimum at a higher frequency, around 300 Hz. Furthermore, the model shows an increase in the ITD threshold when the carrier stimulation rate was increased from 1000 PPS to 5000 PPS (i.e. when $f_m > 16$ Hz) which is not indicated by the behavioral data. The model does not predict well the behavioral data and this is indicated by a low correlation value between model and data of $r = 0.26$ ($p = 0.57$).

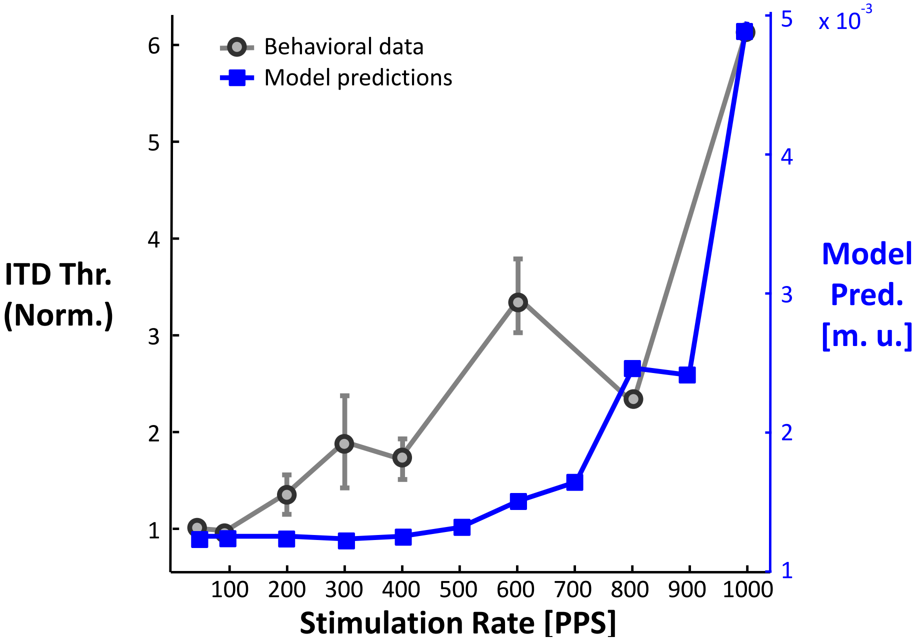

Figure 4.12: Model performance with unmodulated electric pulses. Behavioral data were extracted from six experiments (Egger et al., 2016; Laback et al., 2007; van Hoesel, 2007; van Hoesel and Clark, 1997; van Hoesel et al., 2009; van Hoesel and Tyler, 2003) summarised in Fig. 3 of Laback et al. (2015). The median for each study is estimated across all listeners per study and the resulting means normalised across all studies. The normalised medians are further averaged into a single mean which is shown along with the error bars that indicate the standard error of the mean across the six experiments. This is done to remove any ITD threshold offsets that are imposed by differences in the experimental procedure and does not affect variations due to the stimulation rate. The stimulus duration was 300 ms and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB (which is around 75% chance to generate an action potential for a single maximum current pulse).

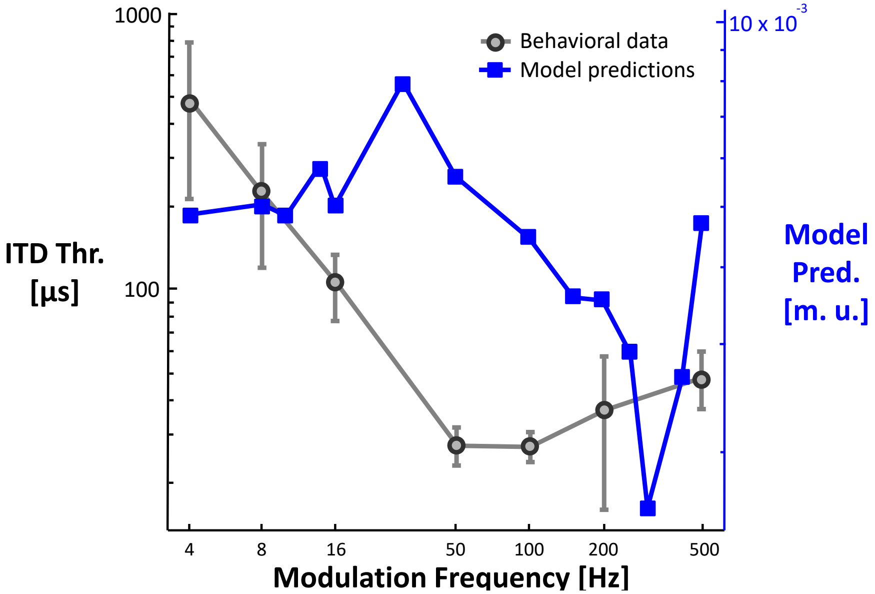

Figure 4.13: Model performance for electric SAM pulses. Experimental data from Noel and Eddington (2013). The stimulus duration was 300 ms and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB, which is around 75% chance of an action potential generated for a single maximum current pulse. Note that for $f_m \leq 16$ Hz the pulse rate was at 1000 PPS, whereas for $f_m > 16$ Hz the pulse rate was at 5000 PPS.

Trapezoidal Modulation

Similarly to the acoustic case of the trapezoidal modulation the modulating envelope is a series of symmetric trapezoids defined by off time, rising and falling slopes and a fixed modulation period of 27.3 Hz. As Fig. 4.14 shows, the behavioral data indicate that as the off time is increased the ITD detection threshold decreases. The model correctly predicts the decrease in ITD threshold as off time increases, with correlation values of $r = 0.88$ ($p = 0.12$), $r = 0.90$ ($p = 0.10$) and $r = 0.91$ ($p = 0.09$) for when the slope is 6, 8 and 12 dBms^{-1} respectively. However, contrary to the acoustic case, the rising and falling slopes do not appear to have significant effect in the ITD detection threshold. This is also well predicted by the model as across the slope correlation has a value of $r = 0.82$ ($p = 0.001$).

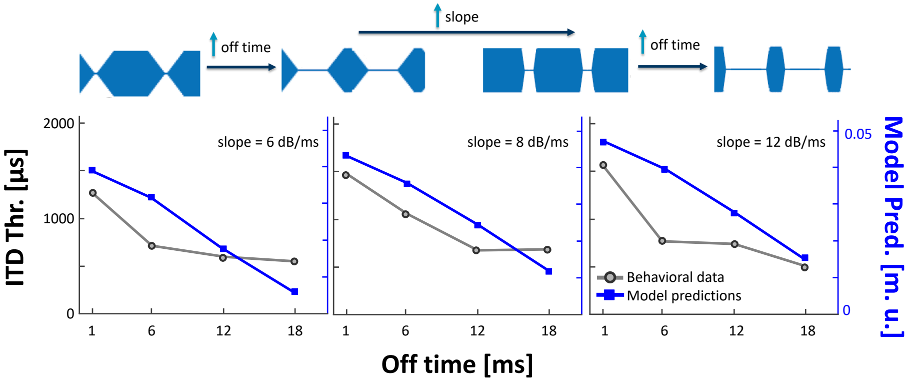

Figure 4.14: Model performance for electric trapezoid modulation pulses. Experimental data from Laback et al. (2011). The stimulus duration was 1 s and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB, which is around 75% chance of an action potential generated for a single maximum current pulse (Sec. 4.3.1).

4.5 Discussion

4.5.1 Model Purpose and Assumptions

We combined acoustic and electric modelling that describe the temporal properties of action potential generation on the AN from both acoustic and electric stimulation with a novel neurometric-psychometric method. Its purpose is to map the temporal pattern of action potentials to the perceptual sensation of ITD (Stüttgen et al., 2011) by estimating the ITD threshold. The binaural ITD sensitivity modelling approach taken here is functional/phenomenological and, as such, a qualitative comparison with behavioral data is feasible. The selection of the various studies providing the behavioral data creates an evaluation framework for the model.

The model presented here is comparable with other models in proposing a generic binaural processor (Breebaart et al., 2001; Dietz et al., 2011; Pulkki and Hirvonen, 2009; Takanen et al., 2014). These models typically consider more complex stimuli with variable ILD and broadband spectrum. However, we wanted to investigate simpler stimuli to reduce the assumptions made about how the central auditory system processes binaural information since there is no paradigm yet that explicitly links human ITD detection performance and ascending neural pathways in the central auditory system. Furthermore, previous correlation-based approaches investigate signals which are continuous representations of periphery (Colburn and Latimer, 1978; Colburn, 1973), whereas here we investigate a discrete representation consisting of binary events. Because a direct application of the short-time interaural cross-correlation (Colburn, 1973) is not applicable, we used the SCC (Joris et al., 2006). This binary events analysis enables the direct comparison of spike trains generated on the AN regardless of the nature of the stimulus, either it being acoustic or electric.

Neurons typically have intrinsic noise in their behavior and, as such, the model output has a degree of variance. In order to minimize this variance of the model output we assumed around 1000 neurons with similar temporal characteristics originating from the periphery and converging to one processing centre which is perfectly binaurally place matched. This assumption could be biophysically unrealistic, especially when considering electrical stimulation, as CI subjects frequently have incomplete neural survival. A further consideration for contrasting the BiCI results and the model prediction is intrinsic variation in behavioral data. This is due to the low number of subjects in the behavioral data, often from a single study. Furthermore, there is typically much unaccounted variability between CI subjects, either due to pathology of the AN and the sound deprived brain, variation in place/locus of stimulating electrodes, or plasticity induced post-operative changes of the tonotopic map.

4.5.2 Performance Assessment and Future Steps

The model performance was tested by comparing its output against behavioral data. For the NH situation, the model predicted the general trends observed for all waveforms tested. However, the model was found to be more sensitive to temporal gaps in stimulus (i.e., off time periods) than the behavioral data indicates. Specifically, this is observed in transposed tones and raised sine tones, where by varying the f_m the model consistently predicts that the minimum f_m for ITD sensitivity is at a lower frequency (i.e., larger temporal gap) than what the behavioral data indicates. This might be caused by the peripheral model's high entrainment to temporal gaps of high intensity stimuli. Further investigations on internal noise, possibly by modelling central processes, can reduce this high entrainment the model exhibits.

The current NH peripheral model uses neurons of a single type of spontaneous activity whereas in reality the AN has a mixture of high spontaneous rate, medium spontaneous rate and low spontaneous rate neurons (Relkin and Doucet, 1991; Rhode and Smith, 1985). The single neuron type was chosen for the sake of simplicity. A preliminary analysis of low spontaneous rate neurons indicated reduced response of neurons to the onset of modulating periods. This affects the model outcome by estimating higher ITD threshold values. Model outcomes when using mixed type neurons, specifically 60% high spontaneous rate, 20% medium spontaneous rate and 20% low spontaneous rate, showed little variation from the reported results which used 100% high spontaneous rate neurons. However, this might be attributed to the high intensity stimuli used in the behavioral studies that form the evaluation framework of the model. Low intensity stimuli could be used to better describe the possible role of low spontaneous rate neurons in ITD detection.

Moreover, the model does not predict the behavioral changes in ITD perception across CF (i.e., different locations on the basilar membrane) reported for SAM and transposed tones. This can be attributed to the fact that no tonotopic considerations were made for the binaural processing of more central auditory pathways. A recent biophysical model which investigated ITD sensitivity was proposed by Chung et al. (2014). They found that their model of brainstem neurons required stronger excitatory synaptic inputs and faster membrane responses for ITD sensitivity at high stimulation rates, whereas weaker excitatory synaptic input and slower membrane responses were necessary at low stimulation rates, for both electric and acoustic stimulation. Their findings suggest the possibility of frequency-dependent differences in the neural mechanisms of binaural processing, which could support the need for a filter, especially for the spectral aspects (i.e., introduce tonotopic considerations) of the proposed model.

For the BiCI situation, the model predicts well the ITD sensitivity for unmodulated pulse trains and temporal variations of trapezoid modulations. However, it does not predict well SAM pulse train ITD sensitivity. This mismatch between model prediction and behavioral data could be attributed to the electric stimulation peripheral model, which is unable to predict the exact variability in action potential generation as a function of pulse rate and firing rate. It is also unable to quantitatively predict sensitivity of AN fibres to small modulation depths (Goldwyn et al., 2012).

A consideration that needs to be made when examining behavioral data for BiCI, especially for SAM pulse trains, is that more data are required, since the data presented are from a single study with 5 subjects (Noel and Eddington, 2013).

Caution needs to be taken when interpreting the correlation values between model and data, especially for the cases where a single condition is examined (e.g. single centre frequency, single trapezoid slope, etc.) because of the low number of data points, typically restricted to 4-6 measurements of variable changes. Across conditions, the number of data points increases to 12-16, based on the number of conditions that were tested.

Further improvement is necessary for the electric stimulation peripheral model. The complete characterization of neuronal responses to electrical stimulation is ongoing (e.g., Boulet et al., 2016). However, this does not directly affect the formulation of the proposed model. Any peripheral model which can produce a sequence of binary events describing the spatiotemporal pattern (i.e., answering the *where* and the *when*) of AN activity could be used in the pipeline of the proposed framework. A more encompassing electric peripheral model is a part of future improvements to this model.

Lastly, although the BCC device was able to output good predictions of the behavioral data trends, it did so analytically using arbitrary “model units”. These are not consistent across different conditions (e.g., a model output of 0.25 could correspond to an ITD threshold of 150 μs in one condition and of 250 μs in another one), making the comparison of the model predictions with the psychoacoustic data more difficult. Ideally, we would like to obtain model predictions in proper time units (μs in this case), which would allow a more straightforward quantitative evaluation of the model performance.

4.6 Conclusions

In summary, the trend of ITD threshold variations caused by frequency changes in *unmodulated* stimuli, (i.e., pure tones and electric pulse trains of clinical stimulation rates) are well described by the model. Specifically, the high frequency cut-off for ITD detection for both acoustic and electric hearing is predicted. Therefore the model can identify the useful region of ITD detection for these types of stimuli. Furthermore, the ITD threshold variation trend caused by modulated stimuli with low frequency modulation (i.e., $\leq$ 30 Hz) is also well described. For mid-high modulation frequencies (i.e., $\geq$ 100 Hz) the model indicates the general trends observed. However it underestimates the behavioral data (i.e. predicts lower thresholds), hence its main strength lies in qualitatively predicting the effect of a change in stimulus parameter on ITD detection, rather than quantitative prediction of threshold in ITD across all possible stimuli.

The applicability of the model proposed here on both BiCI and NH behavioral data support the working hypothesis of a normal operation of the central system for binaural detection, which is also made by Chung et al. (2014). However it should be noted that extended periods of deafness could invalidate this hypothesis. This could cause the general poor ITD sensitivity observed for subjects with extended periods of binaural sensory deprivation (Litovsky et al., 2010).

The model presented here is only suitable for processing narrowband stimuli (i.e., within an ERB). Equivalently for the electrical stimulation case, only a single electrode interaural pair can be considered as there is no explicit modelling of cross-channel interactions. Furthermore, no ILDs were considered for this analysis. These limitations pose no problem when studying static ITD perception of narrowband stimuli, on the contrary they enable fair comparison between model predictions and behavioral data. As such, the psychoacoustic data used in the evaluation framework for model validation used static ITD with narrowband stimuli, or single electrode pair for the electric case, and minimised ILD cues.

CHAPTER 5

Interaural Time Difference Discrimination in Normal and Hearing Impaired Listeners[1]

---

[1] The work presented in this section has been published as Moncada-Torres, A., Joshi, S. N., Prokopiou, A., Wouters, J., Epp, B., and Francart, T. (**2018**). “A framework for computational modelling of interaural time difference discrimination of normal and hearing impaired listeners,” The Journal of the Acoustical Society of America **144**(2), 940–954. Changes are limited to layout, graphical appearance, and minor editing.

5.1 Abstract

Different computational models have been developed to study interaural time difference (ITD) perception. However, only few have used a physiologically-inspired architecture to study ITD discrimination. Furthermore, they do not include aspects of hearing impairment. In this work, a framework was developed to predict ITD thresholds in listeners with normal and impaired hearing. It combines the physiologically-inspired model of the auditory periphery proposed by Zilany et al., [2009, *The Journal of the Acoustical Society of America*, **126**(5), 2390–2412] as a front end with a coincidence detection stage and a neurometric decision device as a back end. It was validated by comparing its predictions against behavioral data for narrowband stimuli from literature. Strong correlations between predictions and data show that the model is able to model ITD discrimination of normal-hearing and hearing-impaired listeners at a group level. Additionally, we used it to explore the effect of different proportions of outer- and inner-hair cell impairment on ITD discrimination.

5.2 Introduction

The best known theory of how we process interaural time differences (ITDs) was proposed by Jeffress (1948). In short, it suggests that ITDs are decoded by neurons in the central auditory system, which receive inputs from both ears and are sensitive to specific delays of their inputs. Thus, they function as coincidence detectors and fire at a particular ITD.

Different computational frameworks based on the Jeffress model have been developed to understand, study, and simulate ITD perception. Initially, binaural phenomena were predicted using cross-correlation of the (continuous) signals arriving to the left and right ears (Sayers and Cherry, 1957). Later on, models started incorporating the correlation of the neural response of both ears at the level of the auditory nerve (AN) as a main component. For instance, Colburn and Latimer (1978); Colburn (1973, 1977) used a coarse model of the AN (consisting of a bandpass filter, automatic gain control, and an exponential rectifier) together with a binaural display to predict ITD discrimination in tone bursts across different levels, as well as binaural unmasking. Stern and Colburn (1978); Stern and Trahiotis (1992), and Stern and Shear (1996) extended this model by including a centrality weighting function (designed to model the larger proportion of coincidence detection units sensitive to shorter ITDs) and a straightness weighting function (designed to emphasize the internal delays consistent across different frequencies). Bernstein and Trahiotis (2012); Lindemann (1986), and Stern and Colburn (1978) also combined ITD with interaural level differences (ILD) information to predict perceptual lateralization as well as binaural detection.

These frameworks have contributed greatly in the studying and understanding of the processing of ITDs by the auditory system. However, their approach for modelling the AN can be considered elemental, since they include a limited amount of aspects regarding its biology. Incorporating the anatomy and the physiological processes underlying the auditory system can provide a better insight of how it processes binaural sounds. More recently, different models have been developed that explicitely include some biological aspects of binaural auditory temporal processing. For instance, Gai et al. (2014) combined a model of the AN with a model of bushy cells to study ITD coding in medial superior olive (MSO) neurons. Takanen et al. (2014) developed a method for visualizing binaural interactions (including sound lateralization using ITDs) using models of the superior olivary complex. Wang et al. (2014) presented a model of the binaural pathway capable of simulating ITD sensitivity of an inferior colliculus (IC) neuron to sinusoidally amplitude modulated tones and broadband noise. However, there are few physiologically-inspired models that have been used to investigate ITD discrimination (e.g., Brughera et al., 2013; Hancock and Delgutte, 2004; Prokopiou et al., 2017).

While substantial progress has been achieved in the study of how normal-hearing (NH) listeners process ITD cues (Bernstein, 2001; Blauert, 1997; Colburn et al., 2006; Grothe et al., 2010; McAlpine, 2005; Stecker and Gallun, 2012; Wang and Brown, 2005), this is not the case for hearing-impaired (HI) listeners (Akeroyd and Whitmer, 2016; Durlach et al., 1981; Moore, 2007). HI listeners frequently present large inter-subject variability (Gabriel et al., 1992; Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; Spencer et al., 2016). Furthermore, it is hard to discern if the HI listeners' binaural hearing ability was affected because of HI itself or because of confounding factors, such as effort, concentration on the task, or age (Gallun et al., 2014; King et al., 2014; Peelle and Wingfield, 2016).

Studying HI using a physiologically-based computational approach (Durlach et al., 1981; Moore, 1996) would allow us to systematically investigate the mechanisms that are detrimental to the (binaural) hearing system and the effect of specific aspects of HI in listeners' performance in different tasks without the confounds of behavioral experiments. Moreover, it could provide us with additional means to improve the design, implementation, and fitting of hearing devices (e.g., hearing aids, cochlear implants). Futhermore, there has been no attempt to model ITD discrimination in HI listeners using a detailed physiologically-based approach.

In this work, we developed a computational framework that adresses these issues. Namely, we used our framework to study ITD discrimination by predicting its thresholds (i.e., just noticeable difference) in both NH *and* HI listeners. It combines the *physiologically-inspired front end* proposed by Zilany et al. (2014, 2009) with a coincidence detection stage and a *neurometric-based decision device as a back end*. We validated it by comparing its ITD threshold predictions for different stimuli against behavioral data reported in literature. Additionally, we used it to explore the effect of different proportions of outer- and inner-hair cell (OHC, IHC) impairment on ITD discrimination.

5.3 Framework Description

The proposed framework is composed of three main building blocks: 1) auditory periphery (responsible for generating spike trains elicited by a given stimulus with an imposed ITD), 2) coincidence detection (its purpose is to compute the imposed ITD), and 3) decision device (its objective is to predict an ITD threshold based on the distributions of the computed ITDs). The framework's block diagram is shown in Fig. 5.1 and is explained in detail below.

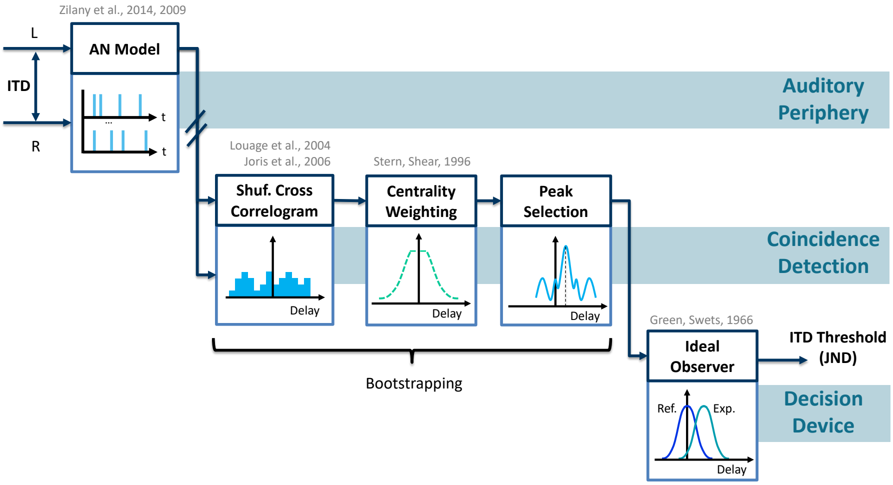

Figure 5.1: Block diagram of the proposed framework. Signals coming from the left and right ear (with an imposed ITD) are processed by the model of the auditory periphery proposed by Zilany et al. (2014, 2009). The resulting spike trains are fed to a coincidence detection stage, which computes the imposed ITD. This process is repeated numerous times in order to generate a distribution of the computations for different ITD values. These distributions are used as an input to the decision device, yielding as an output a predicted threshold value in μs.

5.3.1 Auditory Periphery

As a front end, we chose the model proposed by Zilany et al. (2014, 2009). It is capable of reproducing response properties of AN fibers to acoustic stimuli and has been validated with physiological data over a wider range than previously existing models. Furthermore, it has been successfully used to study a variety of auditory phenomena, such as masking release (Bruce et al., 2013), frequency selectivity (Jennings and Strickland, 2012), neural adaptation to sound level (Zilany and Carney, 2010), sensory responses to musical consonance-dissonance (Bidelman and Heinz, 2011), speech intelligibility in noise (Moncada-Torres et al., 2017), neural coding of chimaeric speech (Heinz and Swaminathan, 2009), and overshoot adaptation (Jennings et al., 2011).

It is composed of different modules, each simulating a particular element of the auditory periphery. For each channel (left and right), the model received as an input an instantaneous pressure waveform with a sampling frequency of 100 kHz. In all cases, the ITD was applied to the right waveform, shifting it in time with respect to the left waveform. For each ear, the stimulus was first passed through a filter emulating the middle ear. Then, the output was passed through a signal path and a control path. The signal path simulated the behavior of the OHC-controlled filtering properties of the basilar membrane in the cochlea and the transduction properties of the IHCs by a succession of non-linear and low-pass filters. The control path simulated the function of the OHCs in controlling basilar membrane filtering. It did so with a wideband basilar membrane filter followed by a non-linearity module and an OHC low-pass filter. The control path output fed back into itself and into the signal path, as well. The IHCs output then went through an IHC-AN synapse module with two power-law adaptation paths (which account for slow and fast adaptations) and a spike generator (which accounts for the activity, adaptation, and refractoriness of the AN response).

The model allows to be tuned with different parameters, which were defined as follows. We chose to simulate human AN fibers with the same characteristic frequency (CF) as the stimulus frequency (in the case of pure tones) or center frequency (in the case of bandpass noise). These were tuned using values reported by Shera et al. (2002). For the sake of simplicity, we decided to simulate only high-spontaneous rate (100 spikes/s) fibers, since in our previous work (Prokopiou et al., 2017) we found that the use of a physiologically relevant mixture of high-, middle-, and low-spontaneous rate fibers had no significant effects on temporal coding at the simulated levels (Table 5.1). We opted for an implementation of the power-law synapse function given it yields a more physiologically accurate response to relatively long stimuli (although at the expense of higher computational power, Zilany et al., 2009). Lastly, we chose to use a pre-determined fixed seed for the fractional Gaussian noise to better separate the (independent) effects of internal (i.e., physiological) and external (i.e., stimulus-driven) noise (Zilany et al., 2014).

Another important parameter was the number of spike trains that needed to be simulated in order to obtain a physiologically relevant response for each ITD computation. We defined this number based on the work by Louage et al. (2006). They recorded spike trains from AN and anteroventral cochlear nucleus neurons of cats to investigate the temporal coding of sound to noise stimuli in afferent neurons and reported collecting ∼3000 spikes per token. Therefore, we made sure that each simulated response had at least 3000 spikes for each channel (left and right). Considering the average number of generated spikes per second to each particular stimulus and the stimulus duration (Sec. 5.4.1), the number of required spike trains ranged from 35 up to 85 (per channel). Thus, we used information of 70–170 neurons for each ITD prediction (2 channels, left and right).

The generation of spike trains is a computationally expensive task, thus we wanted to make the most out of the simulated data by creating new spike train combinations across the simulated neurons. Most importantly, we were interested in obtaining distributions of the ITD predictions (Sec. 5.3.3) without making any *a priori* assumptions about them (Grün, 2009; Ventura, 2010). Bootstrapping is a suitable technique that helped us achieve both. It consists of generating a *data subset* by randomly sampling a *larger data set* with replacement (Witten and Frank, 2011). In our case, the data subset consisted of the required number of spike trains for each condition (35–85 spike trains per channel, Sec. 5.3.1). We defined this number arbitrarily as 20% of the larger data set. Thus, we generated 175–425 spike trains (per channel) in total for each condition.

5.3.2 Coincidence Detection

Shuffled Cross Correlogram

To quantify the temporal pattern of the AN activity, we used the shuffled cross correlogram (SCC, Joris et al., 2006; Louage et al., 2004). The SCC is a modified version of the shuffled auto correlogram (SAC, Joris, 2003). It can be thought as a metric of binaural coincidence detection which compares the timing of the neural firing between the left and right AN fibers. In other words, the SCC quantifies coincident spikes across two different spike trains. The construction of the SCC is illustrated in Fig. 5.2 and is explained as follows.

First, N spike trains from the left ear and N spike trains from the right ear were used as inputs. Then, the forward and backward time intervals between all the spikes of the first left spike train and all the spikes of all the right spike trains were measured. The same was done with all the spikes of the second left spike train and all the spikes of all the right spike trains and so on. These time intervals were tallied into a histogram with binwidth $\Delta\tau$. The latter can be thought as an integration window: if two spikes occur within this time interval, they are considered to be coincident. Louage et al. (2004) originally suggested using a value of 50 μs. However, the resolution of the SCC curve – and of the model's ITD predictions in consequence – is limited by it. Since the objective of the model is to predict ITD thresholds that are behaviorally relevant to human perception, 50 μs was too long. We explored using different $\Delta\tau$ values and found a trade-off between SCC resolution and curve smoothness: smaller $\Delta\tau$ values meant better resolution but more jagged SCC curves (and vice versa). Therefore, we decided on a value of 20 μs.

We were interested in capturing the firing temporal properties of the AN fibers. Thus, the SCC was normalized by the number of spike trains N, the average firing rate r, $\Delta\tau$, and the stimulus duration D. This was done by dividing the SCC by $N^2\,r^2\,\Delta\tau\,D$, making it independent from these parameters and thus dimensionless. In a normalized SCC curve, a count value larger than 1 means that spikes across spike trains tend to be temporally correlated; a count value of 1 shows that there is no (stimulus-induced) temporal correlation; and a count value smaller than 1 indicates anticorrelation (Joris et al., 2006).

Furthermore, we focused our analysis on a window centered at 0 ms spanning ±2 ms (thus 4 ms in total). This is very short compared to the duration of the stimuli used (Sec. 5.4.1). Thus it was not necessary to correct the SCC curve for the distortion caused by the finite duration of the stimulus. An example SCC curve is shown in Fig. 5.3 (panel A).

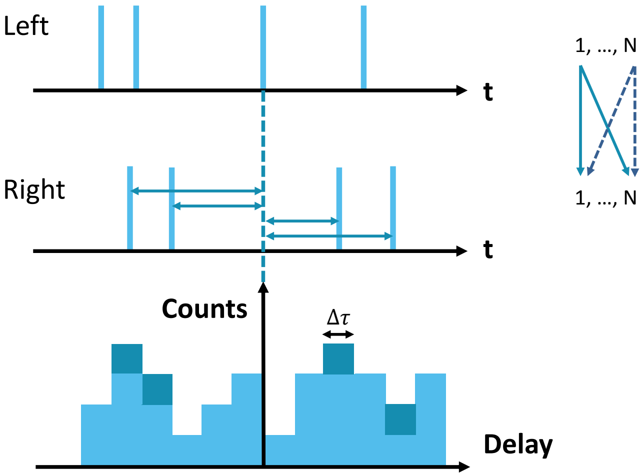

Figure 5.2: Schematic representation of the construction of the Shuffled Cross Correlogram (SCC, Joris et al., 2006; Louage et al., 2004). Notice how the SCC can be plotted as the number of intervals (or counts) versus the delay value. The parameter $\Delta\tau$ (i.e., the histogram bin width) defines the SCC curve and the model's ITD computation resolution. We chose a value of 20 μs.

Centrality Weighting

Different physiological studies of the mammalian auditory system have found that there is a relatively large proportion of neural coincidence detectors that are more sensitive to interaural delays of smaller magnitude (Kuwada et al., 1997, 1987; Kuwada and Yin, 1983; Yin et al., 1986). We accounted for this by introducing a centrality weighting function. The latter served two purposes: to reduce the probability of choosing SCC features for the ITD computation that might be ambiguous (Sec. 5.3.2) and to emphasize the range relevant for human perception.

Several centrality weighting functions have been developed based on psychoacoustic data (e.g., Colburn, 1977; Shackleton et al., 1992; Stern and Shear, 1996). Based on our previous experience, we chose the function proposed by Stern and Shear (1996) yielded the best results. This function p depends on the delay τ and is weakly dependent on the stimulus (center) frequency f, as shown in Eq. 5.1:

$$p(\tau, f) = \begin{cases} 1 & \text{if } |\tau| \leq 200\ \mu\text{s} \\ \frac{e^{-2\,\pi\, k_l(f)\, |\tau|} - e^{-2\,\pi\, k_h\, |\tau|}}{|\tau|} & \text{if } |\tau| > 200\ \mu\text{s} \end{cases} \tag{5.1}$$

with

$$k_l(f) = \begin{cases} 0.1\, f^{1.1} & \text{if } f \leq 1200\text{Hz} \\ 0.1(1200)^{1.1} & \text{if } f > 1200\text{Hz} \end{cases}$$

$$k_h = 3000\,\text{s}^{-1}$$

The generated p spanned across the ±2 ms duration of the SCC curve (Fig. 5.3, panel B). The weighted SCC curve was obtained by multiplying p and the original SCC curve element-wise (Fig. 5.3, panel C).

Peak Selection

We were interested in computing the imposed ITD. We decided to estimate it as the delay value of the maximum peak[2] of the weighted SCC curve, (i.e., the delay value with the largest temporal correlation), as shown in Fig. 5.3 (panel C).

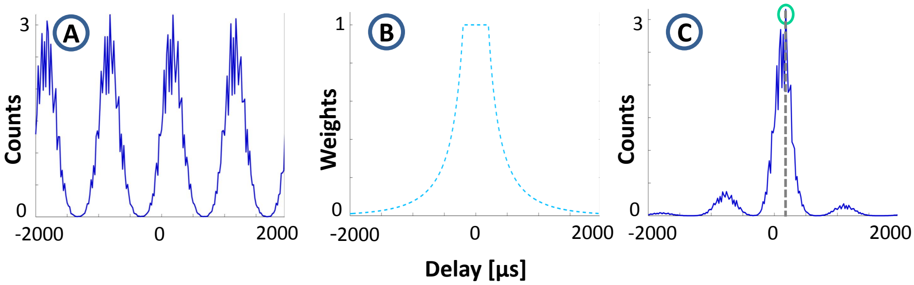

Figure 5.3: Different stages of the Coincidence Detection block for a pure tone with $f = 1$ kHz and an imposed ITD of 160 μs. Panel A shows the raw SCC curve. Panel B shows the centrality weighting function p (adapted from Stern and Shear, 1996). Panel C shows the weighted SCC curve. The ITD was obtained from the latter as the delay value of the maximum peak (dotted line).

---

[2] In their models, Stern and Colburn (1978) and Stern and Shear (1996) estimated the ITD as the centroid of the number of observed coincidences along the delay axis. Bernstein and Trahiotis (2011) found that using the centroid as the decision variable (instead of the maximum peak) was a key factor in their model. However, in our case we explored using the centroid of the SCC curve and reached a different conclusion. Quite often, the SCC curve had a clear peak (which reflected important coincident activity). We found that when using the centroid, this information was lost. The computed ITD distributions of the target cases overlapped quite a lot with the reference one. Additionally, they were much broader (compared to those obtained using the maximum peak). Both of these factors contributed to yielding lower d' values, therefore reducing the slope of the fitted sigmoid and predicting ITD thresholds that were between 200 and 300 μs higher than the behavioral ones. Furthermore, when using the centroid, the model was unable to predict the frequency region of best performance. Therefore, computing the maximum peak was a more suitable choice for our application.

5.3.3 Decision Device

For each case, the Coincidence Detection stage was run 100 times, each time receiving a different set of bootstrapped spike trains (Sec. 5.3.1) and yielding as a result a computed ITD value. Then, we generated distributions of the model computations for different ITDs. Said distributions had a time resolution given by the chosen bin size (20 μs, Sec. 5.3.2). In order to make them continuous and smoother, we fitted a cubic spline using the original data points as anchors. The smoothened distributions were scaled to make sure that the area under the curve added up to the number of ITD computed values (100). Example distributions are shown in Fig. 5.4.

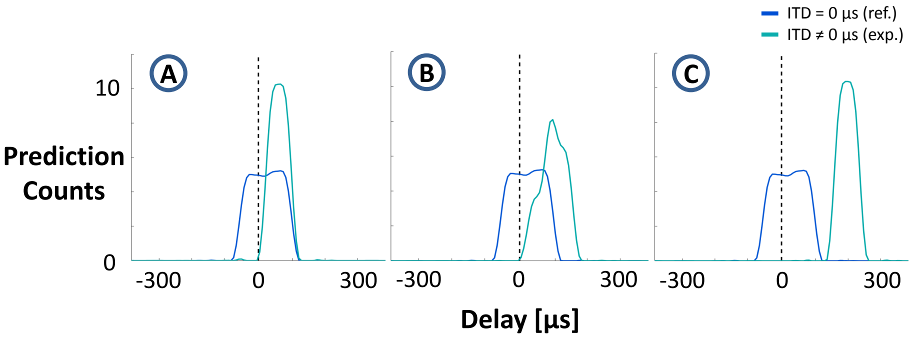

Figure 5.4: Computation distributions for a pure tone with $f = 1$ kHz with imposed ITD values of 40 μs (panel A), 80 μs (panel B), and 160 μs (panel C).

Analogously to a behavioral alternative-forced-choice paradigm, we defined the case where ITD $= 0$ as the *reference condition* and the case where ITD $\neq 0$ as the *experimental condition*. For each experiment, we computed the detection index (d') metric (Green and Swets, 1966), given by Eq. 5.2:

$$d' = \frac{\mu_{\text{ref}} - \mu_{\text{exp}}}{\sqrt{\frac{1}{2}(\sigma^2_{\text{ref}} + \sigma^2_{\text{exp}})}} \tag{5.2}$$

Then, we built a neurometric function by computing d' across different ITD values (namely 10, 20, 40, 80, 160, and 320 μs, Fig. 5.5). Afterwards, we fitted the sigmoid function given by Eq. 5.3 across these points:

$$d' = a + \frac{b - a}{1 + 10^{(c - \text{ITD}) * d}} \tag{5.3}$$

where a is the bottom limit of d'(which we constrained to be not smaller than 0, which translates to a 50% chance of correctly discriminating between the reference and the experimental distribution), b is the top limit of d' (which we constrained to be not larger than 4.65, which translates to a perfect discrimination between the reference and the experimental distribution, Macmillan and Creelman, 2004), c is the ITD_{exp} value that yields a d' value corresponding to 50% of the range of the fitted curve, and d is the slope. The values for these parameters were obtained using a nonlinear regression with iterative least squares estimation. Finally, the ITD threshold was computed as the ITD value corresponding to a point of the fitted sigmoid with a specific d' value. The latter was matched to that of the method used in the behavioral experiment that we wanted to model (Table 5.1). It is worth noting that the predicted ITD threshold is output in ITD units (i.e., μs).

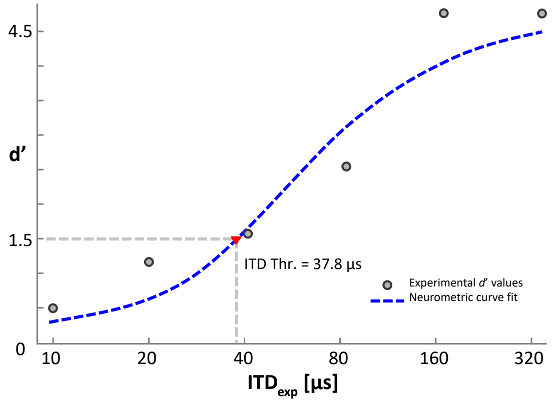

Figure 5.5: Neurometric function for a pure tone with $f = 1$ kHz. In this case, the ITD threshold was computed as the ITD value corresponding to d' of 1.5 (equivalent to 79% of the psychometric function, as used by Brughera et al., 2013), which resulted in a value of 37.8 μs.

5.4 Model Results

5.4.1 Framework Validation

We validated the framework's performance in predicting ITD thresholds by comparing its predictions against behavioral data from different studies reported in literature. We were interested in studies that included NH and HI participants and that used narrow-band stimuli (e.g., pure tones, bandpass noise) with ITDs being the only binaural cue present.

Table 5.1 shows the participants' hearing status, number, and age from the chosen data sets, as well as the characteristics of the stimuli used. In each case, we fed stimuli with these exact same characteristics to the framework. Furthermore, we also generated ITD threshold predictions for additional cases, complementing the behavioral data with computational modelling results. These are shown in Table 5.1 in italics.

Unless stated otherwise, we show the geometric mean and the geometric standard deviation (Kirkwood, 1979) as error bars. In all cases, we show the ITD threshold values on a logarithmic scale (Saberi, 1995).

Table 5.1: Participants and stimuli characteristics of the data sets used in the validation and exploration of the proposed framework.
[1] Brughera et al. (2013); [2] Hawkins and Wightman (1980); [3] Smoski and Trahiotis (1986); [4] Gabriel et al. (1992); [5] Spencer et al. (2016). [6] Model exploration of OHC/IHC loss contribution to HI (Sec. 5.4.2). Conditions where parameters are shown in italics were simulated by the framework, but have no behavioral counterpart. Ref. = reference, av. = average, PT = pure tone, BPN = bandpass noise, thr. = threshold, Dur. = duration, Dec. dev. = decision device.

Ref.	Participants			Stimuli						Dec. dev.
	Hearing status	N	Age [years]	Type	Parameter(s)	Level	Dur. [ms]	Ramps Type	Ramps [ms]	d'
1	NH	5	18 – 22, 31	PT	$f = \{250, 500, 700, 800, 900, 1000, 1200, 1250, 1300, 1350, 1400, \mathit{1450, 1500, 1550}\}$ Hz	70 dB SPL	500	-	100	1.5
2	NH	3	25 (av.)	BPN	$f_c = \{\mathit{250}, 500, \mathit{1000, 2000}, 4000\}$ Hz $bw = \{\mathit{50}, 100, \mathit{200, 300}, 500\}$ Hz	85 dB SPL	250	cos	25	1.3
	HI	8	27 (av.)							
3	NH	2	22, 25	BPN	$f_c = \{\mathit{250}, 500, \mathit{1000, 2000}, 4000\}$ Hz $bw = \{\mathit{50}, 100, \mathit{150, 200}, 270\}$ Hz	80 dB SPL	200	-	25	1.0
	HI	4	20 – 59							
4	NH	2	19, 28	BPN	$f_c = \{250, \mathit{500}, 1000, 2000, \mathit{4000}\}$ Hz $bw = \{58, \mathit{115}, 231, 461, \mathit{919}\}$ Hz	30 dB HL above thr. (Thr. for NH = 0 dB)	300	-	15	1.3
	HI	4	20 – 59		$f_c = \{250, 500, 1000, 2000, 4000\}$ Hz $bw = \{58, 115, 231, 461, 919\}$ Hz					
5	NH	10	19 – 29	BPN	$f_c = \{\mathit{250}, 500, \mathit{1000, 2000}, 4000\}$ Hz $bw = \{\mathit{58}, 115, \mathit{232, 463}, 919\}$ Hz	65 dB SPL	300	linear	15	1.5
	HI	11	20 – 59			thr. $<$ 30 dB: 75 dB SPL 30 dB $<$ thr. $<$ 50 dB: 80 dB SPL thr. $>$ 50 dB: 85 dB SPL				
6	*NH*	-	-	BPN	$f_c = \{\mathit{250, 500, 1000, 2000, 4000}\}$ Hz $bw = \{\mathit{50, 100, 200, 300, 500}\}$ Hz	80 dB SPL	250	cos	25	1.3
	HI	-	-							

NH listeners

Figure 5.6 shows the model's predictions for pure tones in NH listeners. It includes data reported by Brughera et al. (2013). For the latter, we show the median of the ITD threshold for each frequency (together with the original data). When the participants were not able to perform the task, we considered an ITD threshold of $+\infty$ μs. We can see that the predicted ITD threshold values follow the trend of the behavioral data. We further confirmed this by computing Pearson's correlation (ρ) between the behavioral ITD thresholds and the corresponding predictions, yielding a value of $\rho = 0.70$ ($p = 0.02$) across all frequencies and a value of $\rho = 0.89$ ($p < 0.001$) when considering frequencies below 1300 Hz (which is close to the high frequency limit for human ITD discrimination reported by Brughera et al., 2013).

Figure 5.7 shows the model's predictions for bandpass noise in NH listeners. It includes data reported by Gabriel et al. (1992); Hawkins and Wightman (1980); Smoski and Trahiotis (1986) and Spencer et al. (2016). Just like for the pure tones, we can see that overall the framework yields good predictions of the behavioral data. Additionally, the computational modelling predictions suggest that ITD thresholds are the lowest somewhere between 500 and 1000 Hz of the center frequency. Given the low number of participants in each study, we pooled all data together and computed ρ between the behavioral ITD thresholds and their corresponding predictions, yielding a value of $\rho = 0.95$ ($p < 0.001$). Figure 5.10 (panel A) shows the corresponding dispersion plot.

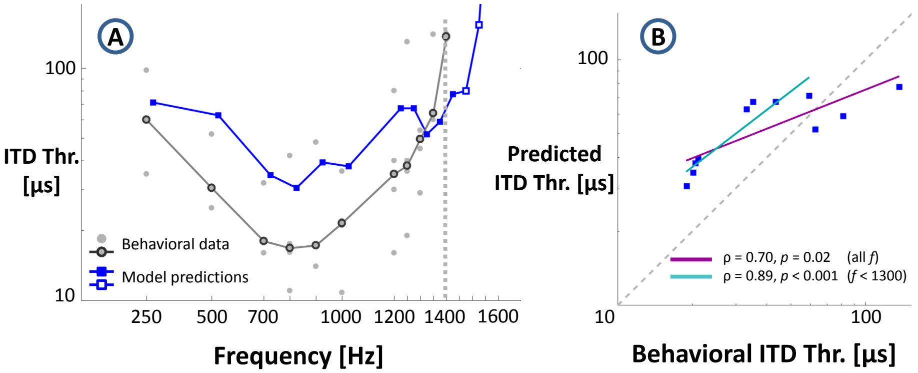

Figure 5.6: Validation of the modelling framework performance for predicting ITD thresholds for pure tones in NH listeners using the behavioral data reported by Brughera et al. (2013) as a reference. Panel A: original data and its median compared with model predictions. The latter correspond to one simulated NH listener (and thus have no error bars). The latter were shifted 25 Hz along the x-axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols. The dotted line represents the highest frequency for human ITD discrimination measured by Brughera et al. (2013). Panel B: dispersion plot of predicted vs behavioral ITD thresholds. The purple line considers data for all frequencies. The blue line considers data for frequencies below 1300 Hz.

HI listeners

One of the most attractive properties of the Zilany et al. (2014, 2009) model is that it is capable of modelling IHC and OHC damage. Physiological studies have revealed that IHC deterioration causes elevation of the AN fiber tuning threshold curves (Liberman and Dodds, 1984), while OHC deterioration additionally causes broadening. Furthermore, OHC damage has also been associated with the reduction in two-tone of AN responses and reduction in the compression of the basilar membrane responses (Liberman, 1984; Liberman and Dodds, 1984; Miller et al., 1997; Robles and Ruggero, 2001; Salvi et al., 1982). The framework incorporates these effects of OHC and IHC impairment using two scaling constants: C_{OHC} and C_{IHC} (Bruce et al., 2003; Zilany and Bruce, 2006), respectively.

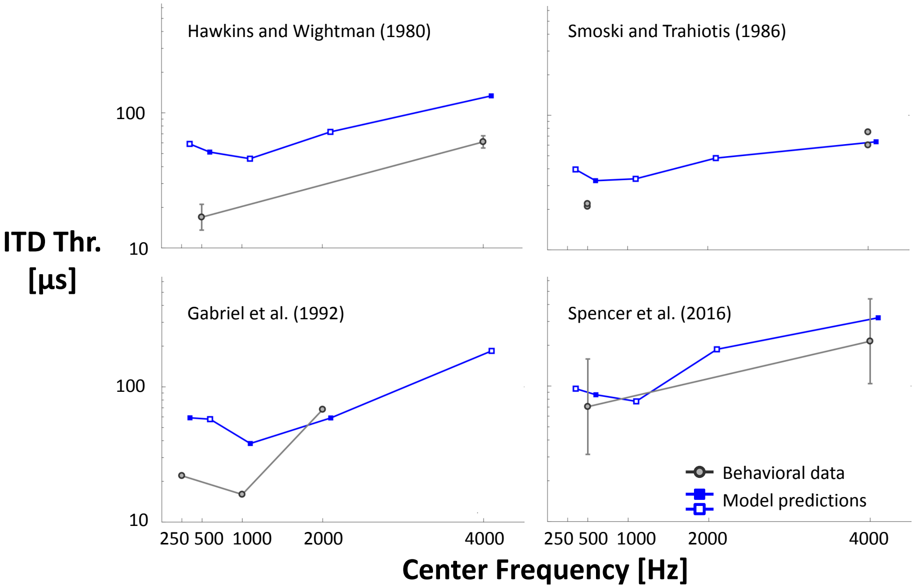

Figure 5.7: Validation of the framework performance for predicting ITD thresholds for bandpass noise in NH listeners using behavioral data reported in different studies as a reference. Error bars denote ± 1 geometric SD across listeners and are reported where available. Model predictions correspond to one simulated NH listener (and thus have no error bars). The latter were shifted 100 Hz along the x-axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols.

C_{OHC} is introduced at the output of the control path. A value of $C_{\mathrm{OHC}} = 1$ represents normal OHC function. This allows for normal behavior of the nonlinear basilar membrane filter, yielding narrow and low tuning thresholds curves and output compression. The closer C_{OHC} gets to 0, the larger the impairment of the OHCs. This modifies the behavior of the basilar membrane filter in two ways: at low sound levels, it increases the tuning curve bandwidth and elevates the thresholds; at moderate to high levels, it reduces (or eliminates) the output compression.

C_{IHC} is introduced in the signal path. Analogous to C_{OHC}, a C_{IHC} value of 1 represents normal IHC function, while values closer to 0 represent larger IHC impairment. This is modeled by lowering the slope of the function that relates basilar membrane vibration to IHC potential, which causes elevated threshold tuning curves.

The model calculated the C_{OHC} and C_{IHC} values for each listener using as an input his/her reported pure-tone audiogram (PTA). It attributed $^2/_3$ of each threshold shift to OHC impairment and the remaining $^1/_3$ to IHC impairment (this proportion is in line with previous studies regarding hearing loss in cats and estimated OHC/IHC detriment in HI listeners in average, Bruce et al., 2003; Lopez-Poveda and Johannesen, 2012; Plack et al., 2004). Specifically, it calculated the listeners' $C_{\text{OHC}}/C_{\text{IHC}}$ by comparing their OHC/IHC threshold shift with a set of previously computed $C_{\text{OHC}}/C_{\text{IHC}}$ values based on the measurements by Shera et al. (2002). A complete description of their computation is given by Bruce et al. (2003) and Zilany and Bruce (2006). Figure 5.8 shows $C_{\text{OHC}}/C_{\text{IHC}}$ values as a function of auditory thresholds for different OHC/IHC proportions and for two CFs (500 and 4000 Hz).

Figure 5.9 shows the model's predictions for bandpass noise in HI listeners. It includes data reported by Gabriel et al. (1992); Hawkins and Wightman (1980); Smoski and Trahiotis (1986); and Spencer et al. (2016). Overall, the framework is capable of following the trends of behavioral data. This is true for data from all studies, except for the case of the Gabriel et al. (1992) dataset, where the model is unable to predict an increase of the ITD threshold at 2000 Hz. Additionally, just like in the NH case, the computational modelling predictions suggest that ITD thresholds are the lowest somewhere between 500 and 1000 Hz of the center frequency. It is worth mentioning that, in most cases, the performance of HI listeners shows a large variability, which is reflected by large error bars. Similarly to the NH case, we pooled all data together and computed ρ between the averaged behavioral ITD thresholds and their corresponding predictions, yielding a value of $\rho = 0.81$ ($p = 0.002$). Figure 5.10 (panel B) shows the corresponding dispersion plot.

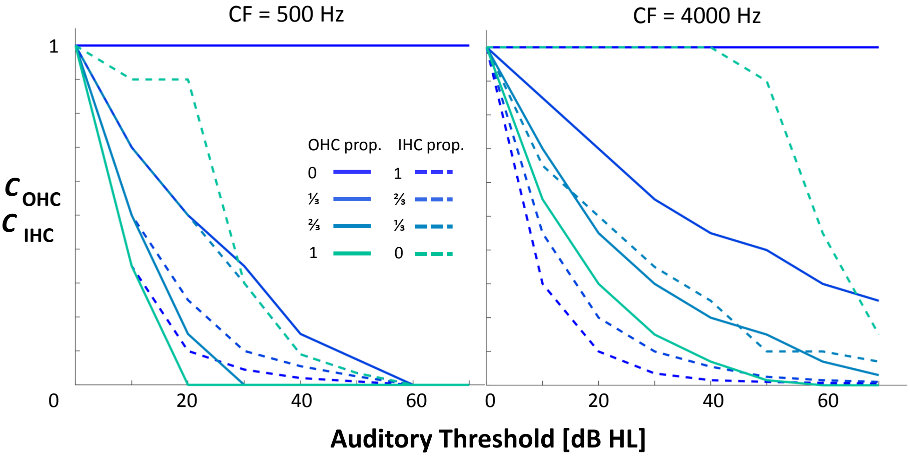

Figure 5.8: C_{OHC} and C_{IHC} values for auditory thresholds for two characteristic frequencies (CF). Different proportions of OHC/IHC impairment (solid/dashed lines) contributing to the threshold shift are shown in different colors.

5.4.2 Contribution of OHC/IHC Loss to HI

Current literature suggests that although both OHC and IHC impairment are responsible for cochlear hearing loss, OHC damage contributes in a larger proportion to the measured threshold shift on average (Bruce et al., 2003; Moore, 2007; Zilany and Bruce, 2006). However, evidence shows that this proportion varies broadly individually across listeners, even among those with identical thresholds (Liberman, 1984; Lopez-Poveda and Johannesen, 2012). Furthermore, anatomical studies have shown that the amount of OHCs and IHCs can vary widely across individuals (Mcgill and Schuknecht, 1976; Wright et al., 1987). Thus if we wish to individualize predictions, more listener-specific factors may have to be taken into account.

Therefore, we explored the effects of different proportions of hair cell loss on ITD discrimination. Besides the original 2:1 proportion attributed to OHC/IHC (Sec. 5.4.1), we investigated the cases when the proportion was 1:0, 1:2, and 0:1. We simulated two HI listeners with a mild and a moderate impairment with flat threshold shifts of 25 and 40 dB hearing level (HL), respectively, across all frequencies (plus a NH listener as a reference). As a stimulus, we used bandpass noise with characteristics defined in the last row of Table 5.1.

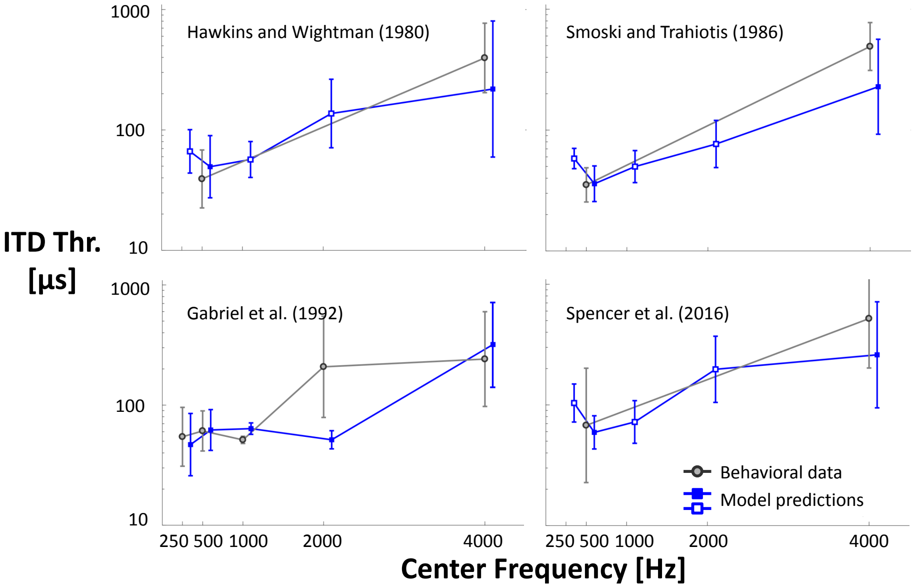

Figure 5.9: Validation of the framework performance for predicting ITD thresholds for bandpass noise in HI listeners using behavioral data reported in different studies as a reference. Error bars denote ± 1 geometric SD across listeners and are reported where available. Model predictions correspond to the simulated HI listeners (i.e., $C_{\mathrm{OHC}} < 1$, $C_{\mathrm{IHC}} < 1$). The latter were shifted 100 Hz along the x-axis for clarity. Model predictions corresponding to behavioral data are shown with filled symbols. Additional model predictions are shown with empty symbols.

These results are shown in Fig. 5.11. On one hand, we can see that in the 25 dB HL case (panel A) there is a slight increase of the thresholds compared to the NH case. However, the curves show little spread, making it hard to see the effect of different OHC/IHC proportions. On the other hand, we can see that in the 40 dB HL case (panel B) not only are the thresholds higher, but the curves have a larger spread. These curves suggest that there is an effect of the different OHC/IHC proportions for frequencies above 500 Hz. Furthermore, it looks like the impairment of OHC has a larger impact on ITD detection: when the auditory threshold shift is completely attributed to the OHCs, the ITD thresholds are the worst (i.e., highest).

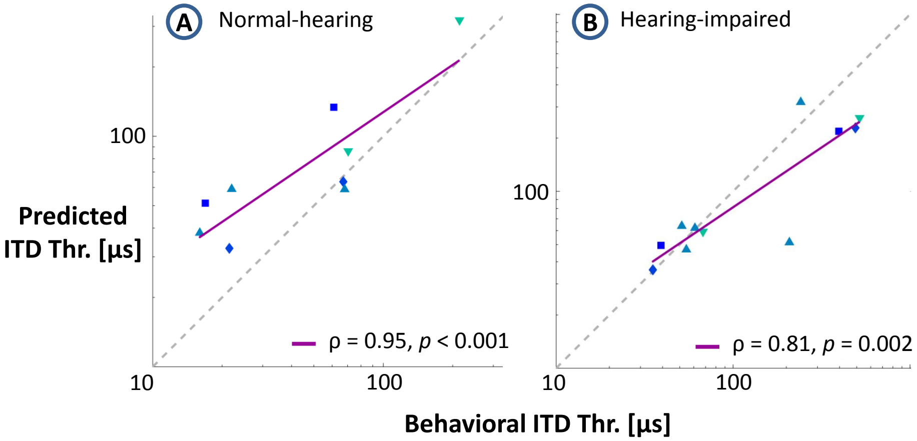

Figure 5.10: Dispersion plots of predicted vs behavioral ITD thresholds for bandpass noise stimuli. Data points from all studies (■ Hawkins and Wightman (1980), ◆ Smoski and Trahiotis (1986), ▲ Gabriel et al. (1992), ▼ Spencer et al. (2016)) were pooled together. In the NH case (panel A), the framework tends to overestimate the ITD thresholds, while in the HI case (panel B), the framework tends to underestimate them.

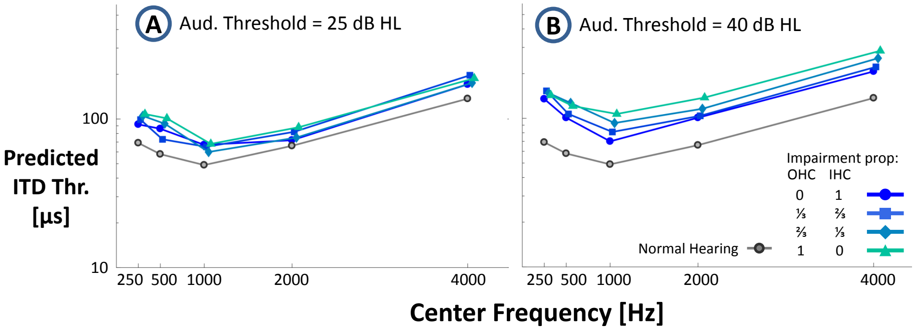

Figure 5.11: Effect of attributing different auditory threshold shift proportions to OHC/IHC impairment for bandpass noise. Its corresponding parameters are given in the last row of Table 5.1.

5.5 Discussion

In this work, we introduced a computational framework that uses a physiologically-inspired model of the AN as a front end and a neurometric decision device as a back end to predict ITD thresholds in NH and HI listeners. We validated its performance by comparing the predicted ITD thresholds for narrow-band stimuli against behavioral data reported in literature. Additionally, we investigated the effect of changing the proportion of impairment attributed to different OHC/IHC combinations.

5.5.1 Framework Structure

The presented computational framework has several advantages. First, the model of the auditory periphery used as a front end (Zilany et al., 2014, 2009, Sec. 5.3.1) is inspired by the anatomy and physiology of the AN and has been validated with a wide range of physiological data. Furthermore, being able to model the individual elements of the auditory pathway allowed us to study them individually. In this particular case, it allowed us to incorporate sensorineural hearing loss as impairment of either the OHCs or the IHCs, something which would not be possible if the representation of the AN was coarser.

Regarding the coincidence detection stage, the SCC (Joris et al., 2006; Louage et al., 2004) basically consists of tallying spike intervals. This process can be thought as a natural display of the information arriving from the auditory periphery (Joris et al., 2005). In other words, it extracts information by comparing temporal coding across contra-lateral AN fibres within an integration window. Our results show that such a relatively simple metric is enough to model ITD discrimination. Additionally, the SCC helped us transform a discrete representation of neural activity (i.e., spike trains) into a continuous representation (i.e., the SCC curve itself), allowing us to further manipulate said representation accordingly (Centrality Weighting) in order to compute an ITD value (Peak Selection).

Finally, the implemented neurometric Decision Device used the distribution of these computations to predict an ITD threshold value. It is worth emphasizing that the latter was given in time units (μs in this case), which made the comparison between the model predictions and the behavioral data intuitive and straightforward. Using the d' metric allowed us to incorporate information about the framework's ITD computations and their dispersion (i.e., variability).

Furthermore, the framework's modular design allowed the use of different building blocks at any of its different stages. For instance, the current Auditory Periphery block could be substituted by any other model that takes a pressure wave as an input and yields spike trains as an output. This is not restricted to models of purely the AN. Models of electrical stimulation could be used to investigate binaural hearing in bilateral cochlear implant users (Prokopiou et al., 2017).

During the framework's development, we were careful to comply with the suggestions proposed by Colburn and Durlach (1978):

1. Relate the model's assumptions and parameters with known physiological data. Our framework included a front end that mimicks the response of the AN based on physiological information.
2. Avoid overfitting of the model's predictions due to excessive parametrization. The framework's parameters were not tuned in a per-case basis. They were defined *a priori* and were the same for all conditions of all simulated experiments (except for the Decision Device criterium [Sec. 5.3.3], which was configured according to the simulated behavioral experiment). Furthermore, we kept the number of parameters to a minimum.
3. Describe quantitatively how the stimuli are processed and how they are affected by (internal) noise. The modular nature of our approach allowed us to examine not only the final output, but also the outputs of the intermediate stages (Sec. 5.3). Additionally, the framework's internal noise (e.g., from the Auditory Periphery, from the Coincidence Detection) was taken into account by the Decision Device to predict the ITD thresholds.
4. Consider perceptual principles when modelling higher, more central portions of the system. The implemented Decision Device (Sec. 5.3.3) models psychophysical aspects of human perception (Green and Swets, 1966).

5. Compare the model predictions with relevant data. We validated the framework with appropriate behavioral datasets previously reported in literature (Sec. 5.4.1).

5.5.2 Framework Validation

The proposed framework was able to predict the trends of behavioral data. In the NH case, we predicted ITD thresholds for pure tones and bandpass noise. For the former, we used the dataset from Brughera et al. (2013) as a reference. Figure 5.6 (panel A) shows that although the framework tended to overestimate the ITD thresholds, it was still able to capture the data trend: low thresholds in the mid-frequency range (700–1000 Hz) and higher thresholds in the low (250–500 Hz) and high (1200–1400 Hz) frequency ranges (with a considerable non-monotonicity at ∼1300 Hz). However, the model does not increase its ITD threshold output as fast as the behavioral data. Additional computational simulations show that the model is only unable to "perform the task" (i.e., output ITD threshold values that were too large) above 1500 Hz. We hypothesized that this was because the auditory periphery stage is unable to capture the loss of phase locking for frequencies above 1300 Hz fast enough. We confirmed this after computing the vector strength (Johnson, 1980) for the spike trains across different frequencies. We obtained vector strength values that slowly decreased monotonically with increasing frequency with an abrupt decrease only at ∼1500 Hz. Therefore, the model finds its performance limit at higher frequencies than we would expect.

In the case of bandpass noise, we used different datasets (Fig. 5.7). The predictions of the datasets from Smoski and Trahiotis (1986) and Spencer et al. (2016) were very close to the reported mean values. The predictions of the datasets from Hawkins and Wightman (1980) and Gabriel et al. (1992)[3] were also able to follow the data trend, although thresholds were overestimated. The model overestimation of the NH thresholds was caused by the SCC resolution. We chose a bin size of 20 μs (smaller values yielded noisy curves leading to incorrect predictions and larger values were too far from human performance). Therefore, the distributions of short experimental ITDs (e.g., 10 or 20 μs) overlapped largely with the distributions of the reference condition (ITD = 0 μs), yielding low

---

[3] This study included only two listeners. However, only the mean was reported, thus we were not able to report the individual data.

d' values. After analysis of different neurometric curves, we found that the points corresponding to these short ITDs were responsible for decreasing the slope of the sigmoid fit of the neurometric function, causing the Decision Device to output higher ITD thresholds. Additionally, the modelling predictions revealed an underlying function that suggests that the best performance (i.e., lowest thresholds) is achieved when the bandpass noise center frequency is between 500 and 1000 Hz. However, this would need to be confirmed with further behavioral measurements.

It is worth mentioning that the model is not able to fully account for the performance detriment between 1000 and 2000 Hz of the Gabriel et al. (1992) data (which is the only study that includes behavioral data for NH listeners at that frequency). The model shows increased thresholds between these two frequencies, but this increase is not as large as in the behavioral case. We attribute this to the model still being sensitive to the temporal fine structure of the stimulus at 2000 Hz. We investigated this by computing the difcor (Louage et al., 2004). The peak of the difcor reflects the temporal fine structure coding strength of the spike trains (and thus of the model). First, we computed the SAC for the left ear spike trains. Then, we computed the SCC between the left and the right ear spike trains. However, the latter corresponded to a polarity-inverted version of the stimulus. Finally, we subtracted them bin by bin. We found that at 1000 Hz, there was strong temporal fine structure coding, with a peak of 3.44. This value decreased only to 2.69 at 2000 Hz. A larger loss of synchrony would result in larger (and thus more accurate) predicted thresholds, as it is in the case at 4000 Hz (with a peak difcor value close to 0).

In the HI case, we predicted ITD thresholds for bandpass noise from the same datasets as in the NH case (Gabriel et al., 1992; Hawkins and Wightman, 1980; Smoski and Trahiotis, 1986; Spencer et al., 2016). Figure 5.9 shows that the framework was able to predict the behavioral trend at a group level. Low center frequency values yielded low ITD thresholds, which the model was able to predict accurately. Increasing center frequency values yielded higher ITD thresholds. The framework was able to capture this trend, although its predictions tended to be lower than the actual behavioral data (in contrast to the NH case). However, these results have to be handled with care, since we cannot affirm that group (i.e., mean) results can be translated to each individual participant (Akeroyd, 2014). High intersubject variability is a common issue in most HI studies, including the ones presented here. This variability could be attributed to a variety of reasons, being low correlations between PTA and listeners' performance in ITD threshold detection tasks one that is commonly suggested (Moore, 2007). We used PTA information to model OHC/IHC impairment (i.e., to compute C_{OHC} and C_{IHC} values, Sec. 5.4.1). We believe this is a good first step towards incorporating HI in our framework. Additionally, just like in the NH case, the modelling predictions suggest that lowest thresholds are achieved when the bandpass noise center frequency is between 500 and 1000 Hz. Likewise, this would need to be confirmed with further behavioral measurements (similar to those performed by Gabriel et al., 1992).

5.5.3 Contribution of OHC/IHC Loss to HI

In the case of the simulated listener with mild impairment (25 dB HL, Fig. 5.11, panel A), we saw that the thresholds were not so different from those of a NH listener. This was expected since thresholds of 25 dB HL are just below the conventional definition of NH of 20 dB HL across all frequencies. Additionally, the stimulus level was well above said threshold.

Of more interest is the case of the simulated listener with moderate impairment (40 dB HL, Fig. 5.11, panel B). These results hint that the impairment of OHCs has a larger impact on the thresholds than impairment of the IHCs. OHC impairment is modelled by the Zilany et al. (2014, 2009) front end as loss of compression, broader tuning, and elevated thresholds (i.e., decreased gain) of the control path (Zilany and Bruce, 2006). These changes in the input-output function as well as in the bandwidth are responsible for the detriment of the temporal coding of the neural fibers. Furthermore, the compression parameters of the model are frequency-dependent (Zhang et al., 2001). The filters' responses are almost linear for frequencies below 500 Hz, while having compressing non-linearities for frequencies above 500 Hz. This explains why we don't see an effect of different OHC/IHC proportions at lower frequencies.

Lopez-Poveda and Johannesen (2012) showed that the OHC/IHC dysfunction contribution to hearing impairment can vary largely across cases, even in listeners with comparable audiometric losses. These results together with the presented model simulations could partially explain the variability observed in the performance of HI listeners in ITD discrimination tasks.

5.5.4 Comparison with Existing Models

Ideally, we would like to quantitatively compare our framework's performance with other similar existing models. Unfortunately, doing such a systematic analysis is not a trivial task (e.g., Ashida et al., 2017; Saremi et al., 2016) and would require a more homogeneous benchmark (Dietz et al., 2018), which is not available for all models. Therefore, further discussion uses a qualitative approach.

First of all, it is worth mentioning that the idea of including physiologically-inspired concepts as components of larger models is not new. For example, Patterson et al. (1995) used the IHC model proposed by Meddis (1988) (which simulates neurotransmitter flow across reservoirs) to convert the basilar membrane motion (coming from either a gammatone or a transmission line filter) into a neural activity pattern. The Zilany et al. (2014, 2009) model has been used as a front end for studying the effect of sensorineural hearing loss on speech intelligibility (Heinz, 2015) and on sound localization in the median plane (Baumgartner et al., 2016). Moreover, very recent work has used that same front end to study ILD perception. The model proposed by Brown and Tollin (2016) subtracted simulated left and right ear AN spike trains within a running temporal window. This simple model was good enough to validate their physiological and psychophysical observations of ILD sensitivity and robustness. Laback et al. (2017) developed a more ellaborate framework. They used the simulated spike trains as an input to an interaural comparison stage where they evaluated the difference in mean discharge rates between left and right ear AN inputs and the variability of these rates over different stimuli presentations. Their model was able to account for a variety of published ILD perception data as well as their own.

Nonetheless, few models have used a physiologically-inspired architecture to study ITD discrimination. To start with, in our previous work (Prokopiou et al., 2017) we investigated binaural temporal discrimination of NH and bilateral cochlear implant users with different stimuli using physiologically-based front ends. However, we did not account for the large proportion of neurons sensitive to smaller ITDs (incorporated in the presented framework by using a centrality weighting function, Sec. 5.3.2). The ITD threshold estimation was done using a novel Binary Classifier Characterisation device. Although the latter yielded good data trend predictions, its approach was more analytical (rather than directly modelling psychophysical procedures). Additionally, the model predictions were given in arbitrary model units (rather than in proper time units [μs], like the current framework), which made the comparison with behavioral data more difficult. Furthermore, we did not investigate the (unaided) HI case.

Hancock and Delgutte (2004) used a neural-pooling model to simulate ITD-sensitive IC neurons to broadband noise. They used gammatone filters to model the processing of the auditory periphery all the way to the IC. Then, they generated a population model which they parametrized with their own physiological data collected in cats. At a single-neuron level, they found very good agreement between physiological data and model predictions of rate-ITD curves. At a population level, they were able to predict ITD thresholds for pure tones and broadband noise. However, using a filterbank to simulate the auditory periphery does not include important details of its physiology and limits is capabilities in further expanding it to include HI at this level. Additionally, they included a so-called "efficiency parameter" (ϵ) for the ITD thresholds predictions. This parameter was empirically adjusted to account for several factors such as stimulus variability, inefficient pooling process, or defficient decision making, but had no physiological relevance. They acknowledged that this approach gave the absolute ITD thresholds predicted values little significance (the comparison with behavioral data was not so straightforward) and thus focused in the trends of ITD threshold changes.

Brughera et al. (2013) minutely investigated ITD threshold discrimination in NH listeners using pure tones across different frequencies, with special focus in the range between 1200 and 1400 Hz. Additionally, they predicted this data set using two different variations of a Hodgkin-Huxley-based MSO neuron model: a lateralization centroid model and a rate-difference model. The former was successful in predicting ITD thresholds at high frequencies, while the latter was successful in predicting ITD thresholds at low and middle frequencies. This rate-difference variation could also predict high frequency thresholds, but only when the model was tuned *ad hoc*. They also proposed using a combination of both types of models in order to predict thresholds across all frequencies. However, neither of these two scenarios (stimulus-specific parameter tuning and/or processing) are desired, since they could very well lead to overfitting.

5.5.5 Framework Limitations

Finally, even though the presented framework was successful in predicting ITD threshold perception trends in NH and HI listeners, there are some shortcomings worth pointing out.

We wanted to investigate the contributions of physiological pre-processing stages shaping the input to the binaural stage. Our results show that using a physiologically-based model of the auditory periphery combined with a coincidence detection stage (such as the SCC) together with a neurometric decision device can be sufficient to model simple binaural processes. In other words, combining information from the auditory periphery with a simple metric of binaural coincidence is enough to model ITD discrimination in NH and HI listeners. Other studies have used a similar approach. For instance, Franken et al. (2014) input cat AN and trapezoid body neural recordings to a "bare bones" model of an MSO neuron. This neuron generated an output spike when its inputs occurred close enough in time. They concluded that the fundamental operation in the mammalian binaural circuit is the coincidence counting of (single) binaural input spikes. This type of simple binaural schemes (either based on metrics or on models) go in line with recent physiological studies that suggest that the ITD sensitivity depends on the exact timing of the excitatory inputs to MSO neurons (van der Heijden et al., 2013). Including a more elaborate MSO stage (such as the models proposed by Brughera et al., 2013; Colburn et al., 2009, 1990; Takanen et al., 2014) could still be benefitial. For example, the MSO has been associated with processing of temporal fine structure, while the lateral superior olive has been associated with processing of the envelope (Remme et al., 2014). Including a two-channel structure (similar to the one used by Dietz et al., 2009) could allow us to further investigate and discern the effect of OHC/IHC loss on the AN responses in terms of temporal fine structure and envelope binaural perception. It could even help reduce the discrepancies between the model predictions and the behavioral data of high-frequency pure tones, since it has been suggested that synchrony loss takes place there (Brughera et al., 2013). Additionally, modelling the MSO neural activity in more detail could help us discern its contribution to auditory brain stem responses (Ashida et al., 2015). Finally, an MSO stage would be crucial to understand better the processing of (sensorineurally-impaired) peripheral input along the binaural auditory pathway.

In the Centrality Weighting block (Sec. 5.3.2) of the Coincidence Detection stage, we used the function proposed by Stern and Shear (1996). This function was obtained using human psychoacoustical data. It would be interesting to explore the effect of centrality-weighting functions obtained using physiological data (e.g., Hancock and Delgutte, 2004; McAlpine et al., 2001). Although these functions were generated using animal data, their use could help in investigating if similar neural distributions exist in humans and to provide a better insight on the underlying assumptions of the psychoacoustic functions.

For the sake of simplicity, we tuned the AN fibers with the same CF as the stimulus frequency (for pure tones) or center frequency (for bandpass noise). Tuning AN fibers to different CFs could allow investigating the effect of cross-frequency integration for ITD perception in NH and HI listeners for broadband stimuli. It could also allow to explore the contribution of off-frequency filters to ITD sensitivity of modulated stimuli (Bernstein and Trahiotis, 2002).

Additionally, we considered ITDs to be the only binaural cue present. However, this is unrealistic, since sounds in a real-world scenario contain ITDs together with ILDs. In order to be able to model this, we would need an additional module that could integrate binaural information across cues. The latter is not a trivial task, since the questions of at what levels of the auditory pathway and to what extent are time and level cues combined are still topics of ongoing research (Brown and Tollin, 2016; Ellinger et al., 2017; Johnson and Hautus, 2010; Palomäki et al., 2005; Phillips and Hall, 2005; Takanen et al., 2014).

Furthermore, the framework fully attributed HI to hair cell damage (Sec. 5.4.1). However, there are additional factors that may affect binaural perception, such as age, cognition, (Gallun et al., 2014; Vercammen et al., 2018) and loss of AN-hair cell synapses (i.e., cochlear synaptopathy, Kujawa and Liberman, 2015; Sergeyenko et al., 2013). The latter could be of special interest, since recently it has been suggested that it could bring ITD thresholds of HI listeners in range of those of NH listeners (Mehraei et al., 2016). To do so, we would need to include low-spontaneous rate fibers (ideally in a physiologically-relevant proportion) and use a front end capable of modelling it (e.g., Verhulst et al., 2018).

Lastly, it is worth mentioning that several assumptions on which the framework's blocks rely on have been validated using animal data. Their translation to human auditory perception still needs to be confirmed by further physiological and psychoacoustical studies (e.g., Salminen et al., 2018).

5.6 Conclusions

We presented a physiologically-based modelling framework capable of predicting ITD thresholds for NH and HI listeners. It uses the model of the AN proposed by Zilany et al. (2014, 2009) as a front end and a Coincidence Detection stage based on the SCC (Joris et al., 2006; Louage et al., 2004) together with a neurometric Decision Device as a back end. The framework was validated by correlating its predictions with behavioral data from literature, yielding ρ values of 0.70 ($p = 0.02$) and 0.95 ($p < 0.001$) for pure tones and bandpass noise in NH listeners, respectively, and of 0.81 ($p = 0.002$) for bandpass noise in HI listeners. These results show that the presented framework is capable of modelling ITD discrimination of NH and HI listeners at a group level. Additionally, we used it to study the contribution of OHC/IHC loss of HI listeners. Model results hint that OHC impairment has a larger impact on ITD discrimination thresholds than damage to the IHCs for frequencies $>$500 Hz.

CHAPTER 6

General Discussion and Conclusion

The general aim of this thesis was to study different dimensions of SI and different auditory processes using computational models. These models exploited the benefits of using a physiologically-based front end (namely the model of the auditory periphery proposed by Zilany et al., 2014, 2009).

This chapter will provide a summary of the core findings of each chapter (Sec. 6.1), discuss practical applications for each of the different frameworks (Sec. 6.2), and give an insight into future research directions (Sec. 6.3). We close the chapter with the general conclusion (Sec. 6.4).

6.1 Core Findings

In Ch. 2, we assessed SI in noise at a word and at phoneme level in NH listeners using two types of objective measures: neurogram-based and filterbank-based.

At a word level, we found that the neurogram-based metrics that incorporated envelope information correlated the strongest with the behavioral scores. These correlations were significantly higher than those found in their temporal fine structure counterparts and than those found using filterbank-based metrics. Our results go in line with literature, which suggests that the envelope component of speech has a larger impact than the temporal fine structure on speech perception (e.g., Drullman, 1995; Shannon et al., 1995; Smith et al., 2002; Swaminathan and Heinz, 2012). A multivariate linear regression analysis also showed that envelope neurogram-based metrics were able to account for a larger proportion of the behavioral data variance, further endorsing their capability for SI prediction.

At a phoneme level, we found that the NSIM metric that incorporated envelope information was the only one that was consistently sensitive to phoneme transitions. This suggests that these transitions have a larger impact on the envelope component of speech (rather than on the temporal fine structure) at the AN level. These results have to be handled with care, though, since currently there is no clear agreement in the literature regarding the influence of phoneme transitions on SI.

Finally, using the presented approach, we could not find evidence that simulating processes at a central level (in the form of the Elhilali et al. (2003) cortical model) provides extra benefit over the information already available at the AN (in the form of the Zilany et al. (2014, 2009) peripheral representation).

In Ch. 3, we studied the effect of level on STMD thresholds in NH listeners. Behavioral experiments showed that at higher ripple densities, STMD thresholds increased (i.e., worsened) with increasing levels. This effect was largest at 4 ripples/oct.

The developed computational model helped us interpret these results, study the contribution of peripheral information to spectrotemporal sensitivity, and to obtain quantitiative threshold predictions. Its results showed that the increased thresholds were caused by the detriment of the spectrotemporal representation at the AN level due to broadening of the cochlear filters. Additionally, higher levels excited a larger number of neurons. The increased neural activity saturated the peripheral representation, diminishing the AN spectrotemporal representation, making it harder to discriminate. The regression analysis revealed that the information at the peripheral level was able to account for a large proportion of the behavioral data variance, supporting its value for the STMD thresholds predictions. We recommend using a fixed presentation level when administering STMD tests, especially when the presentation level cannot be controlled strictly.

In Ch. 4 we presented a framework to predict ITD discrimination thresholds of NH and bilateral cochlear implant listeners. We combined periphery models of acoustic and electric simulation, respectively, with the shuffled cross correlogram (Joris et al., 2006; Louage et al., 2004) and a novel binary classifier characterization. The framework yielded as an output an estimation of the ITD threshold in model units. The predictions of the model were validated with behavioral data from literature.

The presented framework was able to qualitatively predict the effect of different stimulus parameters on ITD detection. The model predicted the data trends for unmodulated stimuli (pure tones in the acoustical case and pulse trains in the electrical case), identifying the frequency regions of best ITD detection performance as well as the high frequency limit. Data trends of low-frequency modulated stimuli were also well described. In the case of mid-high modulation frequencies, the model underestimated the behavioral threshold data.

Finally, in Ch. 5 we presented a framework to predict ITD discrimination thresholds of NH and HI listeners by combining stages of auditory periphery, coincidence detection, and a neurometric decision device. The later allowed the model to yield ITD threshold predictions in relevant (i.e., time) units. Its predictions were validated with behavioral data from literature, as well.

In the NH case, the model was able to capture the data trend, although it had a tendency to overestimate the thresholds. In the HI case, the model was able to predict the behavioral trend at a group level. However, we should keep in mind that group (i.e., mean) results cannot be translated to individual participants, mostly due to the large intersubject variability of HI listeners. In the case of broadband noise (in both NH and HI listeners), additional model simulations revealed an underlying function that suggests that the best performance is achieved when the noise center frequency is between 500 and 1000 Hz.

Additionally, we used the model to study the contribution of different proportions of OHC and IHC loss to ITD discrimination thresholds in HI listeners. We found that a mild impairment ($\sim$25 dB HL) has very little impact on ITD sensitivity. The results of moderate impairment ($\sim$40 dB HL) suggest that the damage to OHCs has a larger impact on ITD thresholds than damage to the IHCs for frequencies >500 Hz. These results could partially explain the variability of HI listeners.

6.2 Applications

The presented frameworks allowed us to gain insight into specific elements of the different auditory processes, to address specific scientific questions, and to predict human performance in different audiological tasks. However, computational models are handy tools that can have many more uses. Thus, we believe it is worth elaborating on their applications under a more pragmatic point of view.

The development of new speech materials can be a very resource-consuming task (e.g., Van Wieringen and Wouters, 2008). First, the speech material itself needs to be selected, constructed, and generated. Afterwards, it is validated behaviorally with listeners. Based on these results, the refinement and tuning of the speech material can go through several iterations. Usually, the behavioral validation of each iteration requires between 10 and 20 new listeners. A computational model like the one presented in Ch. 2 could be a complement to this approach. It could be used in the validation and tuning of the first iterations. Then, a preliminary version of the speech material could be evaluated with actual listeners. This could dramatically reduce the testing time with human participants.

This same framework could also be used to evaluate the effect of speech processing algorithms. The original, clean version of a speech token could be used as the reference, while the token processed by the proposed algorithm could be the "degraded" version. In this case, using a modelling approach could be particularly useful. It could allow extensive parametrization of the speech processing algorithm in a very flexible and scalable manner. This would help to obtain a preliminary version of the algorithm with a set of relevant parameters tuned to have the maximum (simulated) improvement, which could then be tested with actual human listeners.

Current spectral ripple discrimination and spectral/spectrotemporal modulation detection tests present several advantages, such as being strongly correlated with speech perception, providing a language-independent measure of spectral or spectrotemporal sensitivity, and being able to be administered to NH or HI adults or children. However, their use is still not widely spread in clinical scenarios. The framework presented in Ch. 3 could help the development and improvement of this type of tests by further exploring the effect of different parameters (e.g., ripple velocity, ripple density). A better understanding of the mechanisms behind spectral or spectrotemporal sensitivity could bring these tests one step closer to becoming a clinical application.

The frameworks of Ch. 4 and 5 could serve as tools to explore the effects of modifying the stimuli's waveform on ITD sensitivity (e.g., Laback et al., 2011) of a variety of listeners. The output of these models could serve as metric to help generate waveforms that emphasize the coding of binaural temporal cues. Providing the listeners with better ITD cues could potentially improve sound localization as well as speech perception in noise.

These models could also be used as a benchmark to investigate the effect that signal processing schemes have in binaural cues and, therefore, their impact on binaural sensitivity. This is particularly relevant in the bilateral cochlear implantation and bimodal cases, since there is a need to develop strategies that give access to the listeners to binaural (temporal) information (e.g., Francart et al., 2014; Monaghan and Seeber, 2016). This framework could help accelerate the development of such schemes.

6.3 Future Outlook

The work presented in this thesis shows a number of applications of computational physiologically-based models of the auditory system. Not only does it demonstrates that they can be utilized as instruments to answer relevant scientific questions, but it also shows that they are ready to be used practically. In this section, we point to potentially interesting directions of future research and development.

One of the main advantages of the Zilany et al. (2014, 2009) model is that it is able to incorporate the effects of hearing loss due to damage to the OHCs, IHCs, or both. This was something that we exploited when investigating the discrimination of binaural temporal cues in HI listeners in Ch. 5. In a similar fashion, the effect of sensorineural hearing impairment on SI and STMD could be investigated by incorporating it in the frameworks of Ch. 2 and 3, respectively.

One of the things that we kept in mind when guiding the development of the different models was to use a modular approach. This allowed us to analyse the frameworks in a more detailed manner: not only did we examine the final output, but we were also able to look into the outputs of the intermediate stages. A modular architecture could also allow us to use our frameworks as blocks of larger, more complex models. Additionally, it could allow us to plug in either additional blocks to our pipeline or to substitute existing ones with different blocks to explore their effects.

For instance, the “Similarity” block of the model of Ch. 2 could be substituted by either an improved version of the used metrics or by new metrics that process the neurograms using a different approach. Similarly, in the case of Ch. 3, a different method of quantifying the spectrotemporal information at the AN could perhaps provide a better insight of its ability to code that information. In the models of Ch. 4 and 5, we could include additional blocks after the AN corresponding to stages further along the auditory pathway, such as the MSO. In these binaural models, we assumed that the central system of binaural detection is functioning normally in bilateral cochlear implants or HI listeners. In the future, this stage could be substituted by a block that is able to challenge this assumption.

Since we mainly focused on the responses of the AN to different stimuli, this stage (common to all the presented models) is of special interest. Improving the representation of the AN responses could yield better, more relevant results that are closer to the actual physiological processes. Put simply: better models will yield more accurate responses. On this note, it is worth mentioning two very recent models.

Bruce et al. (2018) recently updated the Zilany et al. (2014, 2009) model. In this new version, they improved the stage corresponding to the synapse and spike generation by incorporating a limited number of synaptic sites with adaptive dynamics (as suggested by Peterson and Heil, 2018). This modification improved the model's predictions of a number of published AN data. It would be interesting to re-run the computational simulations presented here with this new version.

Verhulst et al. (2018) published a model of the human auditory periphery capable of simulating human population responses to pure tones and amplitude modulated stimuli in NH and HI listeners using a reduced set of fitting parameters. First, the input stimulus is passed through a middle-ear bandpass filter. Then, it is fed to a transmission-line cochlear model, which was parameterized and validated using human otoacoustic emission recordings. Its output is then passed to a biophysical-inspired representation of the IHCs, a synaptic stage, and three-store diffusion and refractoriness model that generate AN firing rates. This model incorporates hearing impairment as OHC loss by broadening the cochlear filters. Additionally, it is able to incorporate the effect of cochlear synaptopathy (the so called "hidden hearing loss") by changing the number and type of IHC-AN connections. The model was validated using recorded auditory brainstem response and envelope-following response data from NH and HI listeners. It would be attractive to substitute the Zilany et al. (2014, 2009) front end with this AN model and compare the different frameworks' performance.

6.4 General Conclusion

In this thesis, we studied different relevant auditory processes (namely SI, STMD, and ITD sensitivity) using computational models. These models exploited the benefits of having a physiologically-inspired front end in the form of the Zilany et al. (2014, 2009) model of the auditory periphery. The developed frameworks allowed us to answer specific scientific questions regarding the different auditory processes, gain biologically relevant insights into them, and predict human performance of the corresponding audiological tasks.

Computational models are flexible, versatile tools that can function as an ideal complement to the classical anatophysiological and psychoacoustical studies of the auditory system and sound perception. The landscape of the existing models shows that there is no final, universal model that can account for every element, phenomenon, and response of the auditory system. Quite the opposite, the selection of a model will depend on the required level of detail for a specific purpose. Computational models are shared frameworks that are continuously being adapted, reshaped, and improved. Their development is an iterative task that will hardly be ever finished. We believe that the presented models have contributed to the knowledge of the field and hope that they will be studied, exploited, and, most importantly, further improved by the audiological scientific community.

References

Aertsen, A., Johannesma, P. and Hermes, D. **(1980)**. “Spectro-temporal receptive fields of auditory neurons in the grassfrog,” Biological Cybernetics **38**(4), 235–248.

Akeroyd, M. A. **(2014)**. “An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners,” Trends in hearing **18**, 1–7.

Akeroyd, M. A. and Whitmer, W. M. **(2016)**. “Spatial hearing and hearing aids,” *in* “Hearing Aids,” Springer Ch. 7, 181–215.

Allen, J. B. **(2005)**. “Articulation and intelligibility,” Synthesis Lectures on Speech and Audio Processing **1**(1), 1–124.

Anderson, D. J., Rose, J. E., Hind, J. E. and Brugge, J. F. **(1971)**. “Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: frequency and intensity effects,” The Journal of the Acoustical Society of America **49**(4B), 1131–1139.

Anderson, E. S., Oxenham, A. J., Nelson, P. B. and Nelson, D. A. **(2012)**. “Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users.,” The Journal of the Acoustical Society of America **132**(6), 3925–34.

ANSI **(1997)**. *American National Standard: Methods for Calculation of the Speech Intelligibility Index* Vol. 19 Acoustical Society of America New York, USA.

Ashida, G., Funabiki, K. and Kretzberg, J. **(2015)**. “Minimal conductance-based model of auditory coincidence detector neurons,” PloS one **10**(4), 1–16.

Ashida, G., Tollin, D. J. and Kretzberg, J. **(2017)**. “Physiological models of the lateral superior olive,” PLoS computational biology **13**(12), 1–50.

Baumgartner, R., Majdak, P. and Laback, B. (**2016**). “Modeling the effects of sensorineural hearing loss on sound localization in the median plane,” Trends in Hearing **20**, 1–11.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V. and Wellekens, C. (**2007**). “Automatic speech recognition and speech variability: A review,” Speech Communication **49**(10), 763–786.

Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M. and Leek, M. R. (**2013**). “Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners.,” Journal of the American Academy of Audiology **24**(4), 293–306.

Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J. and Lunner, T. (**2016**). “Spectrotemporal Modulation Sensitivity as a Predictor of Speech-Reception Performance in Noise With Hearing Aids.,” Trends in hearing **20**, 1–17.

Bernstein, L. R. (**2001**). “Auditory processing of interaural timing information: new insights,” Journal of neuroscience research **66**(6), 1035–1046.

Bernstein, L. R. and Green, D. M. (**1987**). “The profile-analysis bandwidth,” The Journal of the Acoustical Society of America **81**(6), 1888–1895.

Bernstein, L. R. and Trahiotis, C. (**2002**). “Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”,” The Journal of the Acoustical Society of America **112**(3), 1026–1036.

Bernstein, L. R. and Trahiotis, C. (**2009**). “How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli,” The Journal of the Acoustical Society of America **125**(5), 3234–3242.

Bernstein, L. R. and Trahiotis, C. (**2011**). “Lateralization produced by envelope-based interaural temporal disparities of high-frequency, raised-sine stimuli: Empirical data and modeling,” The Journal of the Acoustical Society of America **129**(3), 1501–1508.

Bernstein, L. R. and Trahiotis, C. (**2012**). “Lateralization produced by interaural temporal and intensitive disparities of high-frequency, raised-sine stimuli: Data and modeling,” The Journal of the Acoustical Society of America **131**(1), 409–415.

Bidelman, G. M. and Heinz, M. G. (**2011**). “Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing,” J. Acoust. Soc. Am. **130**(3), 1488–1502.

Blauert, J. (**1997**). *Spatial Hearing - The Psychophysics of Human Sound Localization* The MIT Press.

Boersma, P. and Weenink, D. (**2014**). “‘Praat: doing phonetics by computer,”’. Version 5.3.16. Date last viewed: 23-10-2014.
URL: *http://www.praat.org/*

Boothroyd, A. and Nittrouer, S. (**1988**). “Mathematical treatment of context effects in phoneme and word recognition,” The Journal of the Acoustical Society of America **84**(1), 101–114.

Boulet, J., White, M. and Bruce, I. C. (**2016**). “Temporal considerations for stimulating spiral ganglion neurons with cochlear implants,” Journal of the Association for Research in Otolaryngology **17**(1), 1–17.

Bradley, J. S. (**1986**). “Predictors of speech intelligibility in rooms,” J. Acoust. Soc. Am. **80**(3), 837–845.

Breebaart, J., Van De Par, S. and Kohlrausch, A. (**2001**). “Binaural processing model based on contralateral inhibition. i. model structure,” The Journal of the Acoustical Society of America **110**(2), 1074–1088.

Briaire, J. J. and Frijns, J. H. (**2000**). “Field patterns in a 3d tapered spiral model of the electrically stimulated cochlea,” Hearing research **148**(1), 18–30.

Bronkhorst, A. W. (**2000**). “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” Acta Acustica united with Acustica **86**(1), 117–128.

Bronkhorst, A. W. and Plomp, R. (**1988**). “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” Acoustical Society of America Journal **83**, 1508–1516.

Brown, A. D. and Tollin, D. J. (**2016**). “Slow temporal integration enables robust neural coding and perception of a cue to sound source location,” Journal of Neuroscience **36**(38), 9908–9921.

Bruce, I. C., Erfani, Y. and Zilany, M. S. (**2018**). “A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites,” Hearing Research **360**, 40–54.

Bruce, I. C., Léger, A. C., Moore, B. C. and Lorenzi, C. (**2013**). “Physiological prediction of masking release for normal-hearing and hearing-impaired listeners,” *in* “Proceedings of Meetings on Acoustics,” Vol. 19 Acoustical Society of America Montreal, Canada pp. 1–8.

Bruce, I. C., Sachs, M. B. and Young, E. D. (**2003**). “An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses,” J. Acoust. Soc. Am. **113**(1), 369–388.

Bruce, I. C., White, M. W., Irlicht, L. S., Leary, S. J. O., Dynes, S., Javel, E. and Clark, G. M. (**1999**). “A stochastic model of the electrically stimulated auditory nerve: single-pulse response,” Biomedical Engineering, IEEE Transactions on **46**(6), 617–629.

Brughera, A., Dunai, L. and Hartmann, W. M. (**2013**). “Human interaural time difference thresholds for sine tones: the high-frequency limit,” The Journal of the Acoustical Society of America **133**(5), 2839–2855.

Carney, L. H. (**1993**). “A model for the responses of low-frequency auditory-nerve fibers in cat,” The Journal of the Acoustical Society of America **93**(1), 401–417.

Carney, L. H. (**1999**). “Temporal response properties of neurons in the auditory pathway.,” Current opinion in neurobiology **9**(4), 442–446.

Cartee, L. A., Miller, C. A. and van den Honert, C. (**2006**). “Spiral ganglion cell site of excitation i: comparison of scala tympani and intrameatal electrode responses,” Hearing research **215**(1-2), 10–21.

Cheng, L. (**2007**). Cochlear-based Transducers: Modeling and Design PhD thesis.

Chi, T., Gao, Y., Guyton, M. C., Ru, P. and Shamma, S. (**1999**). “Spectro-temporal modulation transfer functions and speech intelligibility,” J. Acoust. Soc. Am. **106**(5), 2719–2732.

Choi, J. E., Hong, S. H., Won, J. H., Park, H.-S., Cho, Y. S. Y.-S., Chung, W.-H., Cho, Y. S. Y.-S. and Moon, I. J. (**2016**). “Evaluation of Cochlear Implant Candidates using a Non-linguistic Spectrotemporal Modulation Detection Test.,” Scientific reports **6**, 35235.

Choi, J. E., Won, J. H., Kim, C. H., Cho, Y. S., Hong, S. H. and Moon, I. J. (**2018**). “Relationship between spectrotemporal modulation detection and music perception in normal-hearing, hearing-impaired, and cochlear implant listeners,” Scientific Reports **8**(1), 1–11.

Chung, Y., Delgutte, B. and Colburn, H. S. (**2014**). “Modeling binaural responses in the auditory brainstem to electric stimulation of the auditory nerve,” Journal of the Association for Research in Otolaryngology pp. 1–24.

Colburn, H. and Latimer, J. (**1978**). “Theory of binaural interaction based on auditory-nerve data. iii. joint dependence on interaural time and amplitude differences in discrimination and detection,” The Journal of the Acoustical Society of America **64**(1), 95–106.

Colburn, H. S. (**1973**). “Theory of binaural interaction based on auditory-nerve data. i. general strategy and preliminary results on interaural discrimination,” The Journal of the Acoustical Society of America **54**(6), 1458–1470.

Colburn, H. S. (**1977**). “Theory of binaural interaction based on auditory-nerve data. ii. detection of tones in noise,” The Journal of the Acoustical Society of America **61**(2), 525–533.

Colburn, H. S., Chung, Y., Zhou, Y. and Brughera, A. (**2009**). “Models of brainstem responses to bilateral electrical stimulation,” Journal of the Association for Research in Otolaryngology **10**(1), 91.

Colburn, H. S. and Durlach, N. I. (**1978**). “Models of binaural interaction,” *in* E Carterette and M Friedman, eds, “Handbook of Perception,” Vol. IV (Hearing) Academic Press pp. 467–518.

Colburn, H. S., Shinn-Cunningham, B., Kidd, Jr, G. and Durlach, N. (**2006**). “The perceptual consequences of binaural hearing: Las consecuencias perceptuales de la audición binaural,” International Journal of Audiology **45**(Sup. 1), 34–44.

Colburn, H. S., Yan-an, H. and Culotta, C. P. (**1990**). “Coincidence model of mso responses,” Hearing research **49**(1-3), 335–346.

Cole, R., Yan, Y., Mak, B., Fanty, M. and Bailey, T. (**1996**). “The contribution of consonants versus vowels to word recognition in fluent speech,” *in* “Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on,” Vol. 2 IEEE Atlanta, Georgia, USA pp. 853–856.

Croghan, N. B. H. and Smith, Z. M. (**2018**). “Speech Understanding With Various Maskers in Cochlear-Implant and Simulated Cochlear-Implant Hearing: Effects of Spectral Resolution and Implications for Masking Release,” Trends in Hearing **22**(July), 233121651878727.

Dau, T., Verhey, J. and Kohlrausch, A. (**1999**). “Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers,” J. Acoust. Soc. Am. **106**(5), 2752–2760.

Davies-Venn, E., Nelson, P. and Souza, P. (**2015**). “Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing.,” The Journal of the Acoustical Society of America **138**(1), 492–503.

De Boer, E. (**1975**). “Synthetic whole-nerve action potentials for the cat,” The Journal of the Acoustical Society of America **58**(5), 1030–1045.

De Boer, E. (**1996**). “Mechanics of the cochlea: modeling efforts,” *in* “The cochlea,” Springer pp. 258–317.

De Boer, E. and De Jongh, H. (**1978**). “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,” The Journal of the Acoustical Society of America **63**(1), 115–135.

Deco, G., Scarano, L. and Soto-Faraco, S. (**2007**). “Weber’s law in decision making: integrating behavioral data in humans with a neurophysiological model,” Journal of Neuroscience **27**(42), 11192–11200.

Delgutte, B. (**1996**). “Physiological models for basic auditory percepts,” *in* “Auditory computation,” Springer pp. 157–220.

Deng, L. and Geisler, C. D. (**1987**). “A composite auditory model for processing speech sounds,” The Journal of the Acoustical Society of America **82**(6), 2001–2012.

Dietz, M., Ewert, S. D. and Hohmann, V. (**2009**). “Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities,” The Journal of the Acoustical Society of America **125**(3), 1622–1635.

Dietz, M., Ewert, S. D. and Hohmann, V. (**2011**). “Auditory model based direction estimation of concurrent speakers from binaural signals,” Speech Communication **53**(5), 592–605.

Dietz, M., Lestang, J.-H., Majdak, P., Stern, R. M., Marquardt, T., Ewert, S. D., Hartmann, W. M. and Goodman, D. F. (**2018**). “A framework for testing and comparing binaural models,” Hearing research .

Drullman, R. (**1995**). “Temporal envelope and fine structure cues for speech intelligibility,” J. Acoust. Soc. Am. **97**(1), 585–592.

Duifhuis, H. (**2004**). “Comment on “an approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity”,” The Journal of the Acoustical Society of America **115**(5), 1889–1890.

Durlach, N., Thompson, C. and Colburn, H. (**1981**). “Binaural interaction in impaired listeners: A review of past research,” Audiology **20**(3), 181–211.

Eddins, D. A. and Bero, E. M. (**2007**). “Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region,” The Journal of the Acoustical Society of America **121**(1), 363–372.

Edwards, B. B. W., Struck, C. J., Dharan, P. and Hou, Z. (**1998**). “New Digital Processor for Hearing Loss Compensation based on the Auditory System,” The Hearing Journal **51**(8), 49–52.

Egger, K., Majdak, P. and Laback, B. (**2016**). “Channel interaction and current level affect across-electrode integration of interaural time differences in bilateral cochlear-implant listeners,” Journal of the Association for Research in Otolaryngology **17**(1), 55–67.

Elhilali, M., Chi, T. and Shamma, S. A. (**2003**). “A spectro-temporal modulation index (STMI) for assessment of speech intelligibility,” Speech Commun **41**(2), 331–348.

Ellinger, R. L., Jakien, K. M. and Gallun, F. J. (**2017**). “The role of interaural differences on speech intelligibility in complex multi-talker environments,” The Journal of the Acoustical Society of America **141**(2), EL170–EL176.

Evans, E. (**1972**). “The frequency response and other properties of single fibres in the guinea-pig cochlear nerve,” The Journal of physiology **226**(1), 263–287.

Evans, E. and Palmer, A. R. (**1980**). “Relationship between the dynamic range of cochlear nerve fibres and their spontaneous activity,” Experimental brain research **40**(1), 115–118.

Ewert, S. D. and Dau, T. (**2000**). “Characterizing frequency selectivity for envelope fluctuations,” J. Acoust. Soc. Am. **108**(3), 1181–1196.

Faisal, A. A., Selen, L. P. and Wolpert, D. M. (**2008**). “Noise in the nervous system,” Nature Reviews Neuroscience **9**(4), 292–303.

Field, A., Miles, J. and Field, Z. (**2012**). *Discovering Statistics Using R* SAGE California, USA.

Firszt, J. B., Reeder, R. M. and Skinner, M. W. (**2008**). “Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid,” J Rehabil Res Dev **45**(5), 749–767.

Fogerty, D. and Kewley-Port, D. (**2007**). “Investigating the consonant-vowel boundary: Perceptual contributions to sentence intelligibility,” *in* “Proceedings of Meetings on Acoustics,” Vol. 2 Acoustical Society of America New Orleans, Louisiana, USA p. 060001.

Fogerty, D. and Kewley-Port, D. (**2009**). “Perceptual contributions of the consonant-vowel boundary to sentence intelligibility,” J. Acoust. Soc. Am. **126**(2), 847–857.

Francart, T., Brokx, J. and Wouters, J. (**2009**). "Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation," Journal of the Association for Research in Otolaryngology **10**(1), 131–141.

Francart, T., Lenssen, A. and Wouters, J. (**2011**). "Sensitivity of bimodal listeners to interaural time differences with modulated single- and multiple-channel stimuli," Audiology and Neurotology **16**(2), 82–92.

Francart, T., Lenssen, A. and Wouters, J. (**2014**). "Modulation enhancement in the electrical signal improves perception of interaural time differences with bimodal stimulation," Journal of the Association for Research in Otolaryngology pp. 633–647.

Francart, T., Moonen, M. and Wouters, J. (**2009**). "Automatic testing of speech recognition," Int. J. Audiol. **48**(2), 80–90.

Francart, T., Van Wieringen, A. and Wouters, J. (**2008**). "Apex 3: a multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods **172**(2), 283–293.

Franken, T. P., Bremen, P. and Joris, P. X. (**2014**). "Coincidence detection in the medial superior olive: mechanistic implications of an analysis of input spiking patterns," Frontiers in neural circuits **8**, 42.

French, N. and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**(1), 90–119.

Gabriel, K. J., Koehnke, J. and Colburn, H. S. (**1992**). "Frequency dependence of binaural performance in listeners with impaired binaural hearing," The Journal of the Acoustical Society of America **91**(1), 336–347.

Gai, Y., Kotak, V. C., Sanes, D. H. and Rinzel, J. (**2014**). "On the localization of complex sounds: temporal encoding based on input-slope coincidence detection of envelopes," Journal of neurophysiology **112**(4), 802–813.

Gallun, F. J., Mcmillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M. and Konrad-Martin, D. (**2014**). "Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity," Auditory Cognitive Neuroscience **8**, 172.

Ganong, W. F. (**1980**). "Phonetic categorization in auditory word perception.," Journal of Experimental Psychology: Human Perception and Performance **6**(1), 110.

Geisler, C. D. and Rhode, W. S. (**1982**). "The phases of basilar-membrane vibrations," The Journal of the Acoustical Society of America **71**(5), 1201–1203.

Gelfand, S. A. (**1998**). "Optimizing the reliability of speech recognition scores," J. Speech Lang. Hear. Res. **41**(5), 1088–1102.

Gifford, R. H., Hedley-Williams, A. and Spahr, A. J. (**2014**). "Clinical assessment of spectral modulation detection for adult cochlear implant recipients: A non-language based measure of performance outcomes.," International Journal of Audiology **53**(3), 159–164.

Glasberg, B. R. and Moore, B. C. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hearing research **47**(1), 103–138.

Glasberg, B. R. and Moore, B. C. J. (**2000**). “Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,” Journal of the Acoustical Society of America **108**(5), 2318–2328.

Goldwyn, J. H., Rubinstein, J. T. and Shea-Brown, E. (**2012**). “A point process framework for modeling electrical stimulation of the auditory nerve,” Journal of Neurophysiology **108**(5), 1430–1452.

Green, D. M. and Birdsall, T. G. (**1964**). “Signal detection and recognition by human observers,” Wiley New York, USA.

Green, D. M. and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* John Wiley & Sons.

Greenwood, D. D. (**1961**). “Critical bandwidth and the frequency coordinates of the basilar membrane,” The Journal of the Acoustical Society of America **33**(10), 1344–1356.

Greenwood, D. D. (**1990**). “A cochlear frequency-position function for several species – 29 years later,” The Journal of the Acoustical Society of America **87**(6), 2592–2605.

Grothe, B., Pecka, M. and McAlpine, D. (**2010**). “Mechanisms of sound localization in mammals,” Physiological reviews **90**(3), 983–1012.

Grün, S. (**2009**). “Data-driven significance estimation for precise spike correlation,” Journal of Neurophysiology **101**(3), 1126–1140.

Hancock, K. E. and Delgutte, B. (**2004**). “A physiologically based model of interaural time difference discrimination,” The Journal of neuroscience **24**(32), 7110–7117.

Hawkins, D. B. and Wightman, F. L. (**1980**). “Interaural time discrimination ability of listeners with sensorineural hearing loss,” Audiology **19**(6), 495–507.

Heinz, M. G. (**2015**). “Neural modelling to relate individual differences in physiological and perceptual responses with sensorineural hearing loss,” *in* “Proceedings of the International Symposium on Auditory and Audiological Research,” Vol. 5 pp. 137–148.

Heinz, M. G. and Swaminathan, J. (**2009**). “Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech,” J. Assoc. Res. Otolaryngol. **10**(3), 407–423.

Hellman, R. P. (**1999**). “Cross-modality matching: A tool for measuring loudness in sensorineural impairment,” Ear and hearing **20**(3), 193–213.

Hines, A. and Harte, N. (**2010**). “Speech intelligibility from image processing,” Speech Commun **52**(9), 736–752.

Hines, A. and Harte, N. (**2012**). “Speech intelligibility prediction using a neurogram similarity index measure,” Speech Commun **54**(2), 306–320.

Holmes, S. D., Sumner, C. J., O'Mard, L. P. and Meddis, R. (**2004**). "The temporal representation of speech in a nonlinear model of the guinea pig cochlea," The Journal of the Acoustical Society of America **116**(6), 3534–3545.

Hornsby, B. W. (**2004**). "The speech intelligibility index: What is it and what's it good for?," The Hearing Journal **57**(10), 10–17.

Hossain, M. E., Jassim, W. A. and Zilany, M. S. (**2016**). "Reference-free assessment of speech intelligibility using bispectrum of an auditory neurogram," PloS one **11**(3), e0150415.

Hubbard, A. E. and Mountain, D. C. (**1996**). "Analysis and synthesis of cochlear mechanical function using models," *in* "Auditory Computation," Springer pp. 62–120.

Irino, T. and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter: The gammachirp," The Journal of the Acoustical Society of America **101**(1), 412–419.

Irino, T. and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," The Journal of the Acoustical Society of America **109**(5), 2008–2022.

Javel, E., Geisler, C. D. and Ravindran, A. (**1978**). "Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analyses," The Journal of the Acoustical Society of America **63**(4), 1093–1104.

Jeffress, L. A. (**1948**). "A place theory of sound localization.," Journal of Comparative and Physiological Psychology **41**(1), 35.

Jenkins, J. J., Strange, W. and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**(2), 1030–1043.

Jennings, S. G., Heinz, M. G. and Strickland, E. A. (**2011**). "Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot," J. Assoc. Res. Otolaryngol. **12**(3), 345–360.

Jennings, S. G. and Strickland, E. A. (**2012**). "Evaluating the effects of olivocochlear feedback on psychophysical measures of frequency selectivity," J. Acoust. Soc. Am. **132**(4), 2483–2496.

Johnson, B. W. and Hautus, M. J. (**2010**). "Processing of binaural spatial information in human auditory cortex: neuromagnetic responses to interaural timing and level differences," Neuropsychologia **48**(9), 2610–2619.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," The Journal of the Acoustical Society of America **68**(4), 1115–1122.

Jørgensen, S. and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Soc. Am. **130**(3), 1475–1487.

Jørgensen, S., Ewert, S. D. and Dau, T. (**2013**). “A multi-resolution envelope-power based model for speech intelligibility,” J. Acoust. Soc. Am. **134**(1), 436–446.

Joris, P. X. (**2003**). “Interaural time sensitivity dominated by cochlea-induced envelope patterns,” The Journal of neuroscience **23**(15), 6345–6350.

Joris, P. X., Louage, D. H., Cardoen, L. and van der Heijden, M. (**2006**). “Correlation index: a new metric to quantify temporal coding,” Hearing research **216**, 19–30.

Joris, P. X., van der Heijden, M., Louage, D. H., Van de Sande, B. and Van Kerckhoven, C. (**2005**). “Dependence of binaural and cochlear “best delays” on characteristic frequency,” *in* “Auditory Signal Processing,” Springer pp. 477–483.

Kiang, N. Y.-S. (**1965**). “Discharge patterns of single fibers in the cat’s auditory nerve,” Technical report Massachusetts Institute of Technology - Cambridge Research Lab of Electronics.

Kidd, G., Mason, C. R., Best, V. and Marrone, N. (**2010**). “Stimulus factors influencing spatial release from speech-on-speech masking,” Acoustical Society of America **128**(4), 1965–1978.

Kim, S.-G., Richter, W. and Uğurbil, K. (**1997**). “Limitations of temporal resolution in functional mri,” Magnetic resonance in medicine **37**(4), 631–636.

King, A., Hopkins, K. and Plack, C. J. (**2014**). “The effects of age and hearing loss on interaural phase difference discrimination a,” The Journal of the Acoustical Society of America **135**(1), 342–351.

Kirkwood, T. (**1979**). “Geometric mean and measures of dispersion,” Biometrics **19**, 908–909.

Kistler, D. J. and Wightman, F. L. (**1992**). “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” The Journal of the Acoustical Society of America **91**(3), 1637–1647.

Kowalski, N., Depireux, D. A. and Shamma, S. A. (**1996**). “Analysis of dynamic spectra in ferret primary auditory cortex. i. characteristics of single-unit responses to moving ripple spectra,” Journal of neurophysiology **76**(5), 3503–3523.

Kryter, K. D. (**1962**). “Methods for the calculation and use of the articulation index,” J. Acoust. Soc. Am. **34**(11), 1689–1697.

Kujawa, S. G. and Liberman, M. C. (**2015**). “Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss,” Hearing research **330**, 191–199.

Kulkarni, A. and Colburn, H. S. (**2004**). “Infinite-impulse-response models of the head-related transfer function,” The Journal of the Acoustical Society of America **115**(4), 1714–1728.

Kuwada, S., Batra, R. and Fitzpatrick, D. C. (**1997**). “Neural processing of binaural temporal cues,” Binaural and spatial hearing in real and virtual environments pp. 399–425.

Kuwada, S., Stanford, T. R. and Batra, R. (**1987**). “Interaural phase-sensitive units in the inferior colliculus of the unanesthetized rabbit: effects of changing frequency,” Journal of Neurophysiology **57**(5), 1338–1360.

Kuwada, S. and Yin, T. C. (**1983**). “Binaural interaction in low-frequency neurons in inferior colliculus of the cat. i. effects of long interaural delays, intensity, and repetition rate on interaural delay function,” Journal of Neurophysiology **50**(4), 981–999.

Laback, B., Dietz, M. and Joris, P. (**2017**). “Temporal effects in interaural and sequential level difference perception,” The Journal of the Acoustical Society of America **142**(5), 3267–3283.

Laback, B., Egger, K. and Majdak, P. (**2015**). “Perception and coding of interaural time differences with bilateral cochlear implants,” Hearing research **322**, 138–150.

Laback, B., Majdak, P. and Baumgartner, W.-D. (**2007**). “Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing,” J Acoust Soc Am **121**(4), 2182.

Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W.-D. and Pok, S.-M. (**2011**). “Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearinga),” The Journal of the Acoustical Society of America **130**(3), 1515–1529.

Langner, F., Saoji, A. A., Büchner, A. and Nogueira, W. (**2017**). “Adding simultaneous stimulating channels to reduce power consumption in cochlear implants,” Hearing Research **345**, 96–107.

Lee, J. H. and Kewley-Port, D. (**2009**). “Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners,” J. Acoust. Soc. Am. **125**(2), 1153–1163.

Leek, M. R. and Summers, V. (**1996**). “Reduced frequency selectivity and the preservation of spectral contrast in noise,” The Journal of the Acoustical Society of America **100**(3), 1796–1806.

Levitt, H. (**1971**). “Transformed up-down methods in psychoacoustics,” The Journal of the Acoustical society of America **49**(2B), 467–477.

Liberman, M. C. (**1978**). “Auditory-nerve response from cats raised in a low-noise chamber,” The Journal of the Acoustical Society of America **63**(2), 442–455.

Liberman, M. C. (**1982*a***). “The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency,” The Journal of the Acoustical Society of America **72**(5), 1441–1449.

Liberman, M. C. (**1982*b***). “Single-neuron labeling in the cat auditory nerve,” Science **216**(4551), 1239–1241.

Liberman, M. C. (**1984**). “Single-neuron labeling and chronic cochlear pathology. i. threshold shift and characteristic-frequency shift,” Hearing research **16**(1), 33–41.

Liberman, M. C. and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. iii. stereocilia damage and alterations of threshold tuning curves," Hearing research **16**(1), 55–74.

Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals," The Journal of the Acoustical Society of America **80**(6), 1608–1622.

Litovsky, R. Y., Jones, G. L., Agrawal, S. and van Hoesel, R. (**2010**). "Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans.," J Acoust Soc Am **127**(1), 400–414.

Litvak, L. M., Spahr, A. J., Saoji, A. A. and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners.," The Journal of the Acoustical Society of America **122**(2), 982–991.

Lopez-Poveda, E. A. (**2005**). "Spectral processing by the peripheral auditory system: facts and models," International review of neurobiology **70**, 7–48.

Lopez-Poveda, E. A. and Eustaquio-Martín, A. (**2006**). "A biophysical model of the inner hair cell: the contribution of potassium currents to peripheral auditory compression," Journal of the Association for Research in Otolaryngology **7**(3), 218–235.

Lopez-Poveda, E. A. and Johannesen, P. T. (**2012**). "Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," Journal of the Association for Research in Otolaryngology **13**(4), 485–504.

Lopez-Poveda, E. A. and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," The Journal of the Acoustical Society of America **100**(5), 3248–3259.

Lopez-Poveda, E., O'Mard, L. and Meddis, R. (**1998**). "A revised computational inner hair cell model," *in* "Psychophysical and Physiological Advances in Hearing," Whurr, London pp. 102–108.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S. and Moore, B. C. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proceedings of the National Academy of Sciences **103**(49), 18866–18869.

Louage, D. H., Joris, P. X. and van der Heijden, M. (**2006**). "Decorrelation sensitivity of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise," The Journal of neuroscience **26**(1), 96–108.

Louage, D. H., van der Heijden, M. and Joris, P. X. (**2004**). "Temporal properties of responses to broadband noise in the auditory nerve," Journal of Neurophysiology **91**(5), 2051–2065.

Lyon, R. F. (**1997**). "All-pole models of auditory filtering," Diversity in auditory mechanics pp. 205–211.

Lyon, R. and Shamma, S. (**1996**). "Auditory representations of timbre and pitch," *in* "Auditory computation," Springer New York, NY, USA pp. 221–270.

Macmillan, N. A. and Creelman, C. D. **(2004)**. *Detection theory: A user's guide* Psychology press.

Macpherson, E. A. and Middlebrooks, J. C. **(2002)**. "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited.," J Acoust Soc Am **111**(5 Pt 1), 2219–2236.

Mamun, N., Jassim, W. A. and Zilany, M. S. **(2015)**. "Prediction of speech intelligibility using a neurogram orthogonal polynomial measure (nopm)," IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(4), 760–773.

McAlpine, D. **(2005)**. "Creating a sense of auditory space," The Journal of physiology **566**(1), 21–28.

McAlpine, D., Jiang, D. and Palmer, A. R. **(2001)**. "A neural code for low-frequency sound localization in mammals," Nature neuroscience **4**(4), 396.

Mcgill, T. J. and Schuknecht, H. F. **(1976)**. "Human cochlear changes in noise induced hearing loss," The Laryngoscope **86**(9), 1293–1302.

McNeal, D. R. **(1976)**. "Analysis of a model for excitation of myelinated nerve," Biomedical Engineering, IEEE Transactions on (4), 329–337.

Meddis, R. **(1986)**. "Simulation of mechanical to neural transduction in the auditory receptor," The Journal of the Acoustical Society of America **79**(3), 702–711.

Meddis, R. **(1988)**. "Simulation of auditory–neural transduction: Further studies," The Journal of the Acoustical Society of America **83**(3), 1056–1063.

Meddis, R., Lopez-Poveda, E., Fay, R. R. and Popper, A. N. **(2010)**. *Computational models of the auditory system* Vol. 35 Springer.

Mehraei, G., Gallun, F. J., Leek, M. R. and Bernstein, J. G. W. **(2014)**. "Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility," The Journal of the Acoustical Society of America **136**(1), 301–316.

Mehraei, G., Hickox, A. E., Bharadwaj, H. M., Goldberg, H., Verhulst, S., Liberman, M. C. and Shinn-Cunningham, B. G. **(2016)**. "Auditory brainstem response latency in noise as a marker of cochlear synaptopathy," Journal of Neuroscience **36**(13), 3755–3764.

Miller, C. A., Robinson, B. K., Rubinstein, J. T., Abbas, P. J. and Runge-Samuelson, C. L. **(2001)**. "Auditory nerve responses to monophasic and biphasic electric stimuli," Hearing Research **151**(1-2), 79–94.

Miller, N. **(2013)**. "Measuring up to speech intelligibility," International Journal of Language & Communication Disorders **48**(6), 601–612.

Miller, R. L., Schilling, J. R., Franck, K. R. and Young, E. D. **(1997)**. "Effects of acoustic trauma on the representation of the vowel/ε/in cat auditory nerve fibers," The Journal of the Acoustical Society of America **101**(6), 3602–3616.

Mino, H., Rubinstein, J. T. and White, J. A. (**2002**). “Comparison of algorithms for the simulation of action potentials with stochastic sodium channels,” Annals of biomedical engineering **30**(4), 578–587.

Møller, A. R. (**2000**). “Hearing: its physiology and pathophysiology,”.

Monaghan, J. J. and Seeber, B. U. (**2016**). “A method to enhance the use of interaural time differences for cochlear implants in reverberant environments,” The Journal of the Acoustical Society of America **140**(2), 1116–1129.

Moncada-Torres, A., van Wieringen, A., Bruce, I. C., Wouters, J. and Francart, T. (**2017**). “Predicting phoneme and word recognition in noise using a computational model of the auditory periphery,” The Journal of the Acoustical Society of America **141**(1), 300–312.

Moore, B. C. (**1996**). “Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids,” Ear and hearing **17**(2), 133–161.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological, and Technical Issues* Wiley Series in Human Communication Science.

Moore, B. C. J. (**2013**). *An Introduction to the Psychology of Hearing* sixth edition edn Brill.

Moore, B. C. J. and Glasberg, B. R. (**1987**). “Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,” Hearing research **28**(2-3), 209–225.

Motz, H. and Rattay, F. (**1986**). “A study of the application of the Hodgkin-Huxley and the Frankenhaeuser-Huxley model for electrostimulation of the acoustic nerve,” Neuroscience **18**(3), 699–712.

Mountain, D. C. and Hubbard, A. E. (**1996**). “Computational analysis of hair cell and auditory nerve processes,” *in* “Auditory computation,” Springer pp. 121–156.

Müller, M. and Robertson, D. (**1991**). “Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibres in the guinea pig,” Hearing research **57**(1), 63–70.

Müsch, H. and Buus, S. (**2001**). “Using statistical decision theory to predict speech intelligibility. i. model structure,” J. Acoust. Soc. Am. **109**(6), 2896–2909.

Nicoletti, M., Rudnicki, M. and Hemmert, W. (**2010**). “A model of the auditory nerve for acoustic- and electric excitation,” Frontiers in Computational Neuroscience **4**(0), 5.

Nicoletti, M., Wirtz, C. and Hemmert, W. (**2013**). “Modeling sound localization with cochlear implants,” *in* “The technology of binaural listening,” Springer pp. 309–331.

Noel, V. A. and Eddington, D. K. (**2013**). “Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences a,” The Journal of the Acoustical Society of America **133**(4), 2314–2328.

Offeciers, E., Morera, C., Müller, J., Huarte, A., Shallop, J. and Cavalle, L. (**2005**). "International consensus on bilateral cochlear implants and bimodal stimulation: Second meeting consensus on auditory implants, 19-21 February 2004, Valencia, Spain," Acta Oto-laryngologica **125**(9), 918–919.

Palmer, A. and Russell, I. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hearing research **24**(1), 1–15.

Palomäki, K. J., Tiitinen, H., Mäkinen, V., May, P. J. and Alku, P. (**2005**). "Spatial processing in human auditory cortex: the effects of 3d, itd, and ild stimulation techniques," Cognitive brain research **24**(3), 364–379.

Patterson, R. D., Allerhand, M. H. and Giguere, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," The Journal of the Acoustical Society of America **98**(4), 1890–1894.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**(2), 413–422.

Peelle, J. E. and Wingfield, A. (**2016**). "The neural consequences of age-related hearing loss," Trends in Neurosciences **39**(7), 486–497.

Peterson, A. J. and Heil, P. (**2018**). "A simple model of the inner-hair-cell ribbon synapse accounts for mammalian auditory-nerve-fiber spontaneous spike times," Hearing research **363**, 1–27.

Phillips, D. P. and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hearing research **202**(1), 188–199.

Plack, C. J., Drga, V. and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," The Journal of the Acoustical Society of America **115**(4), 1684–1695.

Prokopiou, A., Moncada-Torres, A., Wouters, J. and Francart, T. (**2017**). "Functional modelling of interaural time difference discrimination in acoustical and electrical hearing," Journal of Neural Engineering **14**(4), 1–21.

Pulkki, V. and Hirvonen, T. (**2009**). "Functional count-comparison model for binaural decoding," Acta Acustica united with Acustica **95**(5), 883–900.

Rattay, F. (**1986**). "Analysis of Models for External Stimulation of Axons," IEEE Transactions on Biomedical Engineering **33**(10), 974–977.

Relkin, E. M. and Doucet, J. R. (**1991**). "Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons," Hearing Research **55**(2), 215–222.

Remme, M. W., Donato, R., Mikiel-Hunter, J., Ballestero, J. A., Foster, S., Rinzel, J. and McAlpine, D. (**2014**). "Subthreshold resonance properties contribute to the efficient coding of auditory spatial cues," Proceedings of the National Academy of Sciences **111**(22), E2339–E2348.

Rhode, W. S. and Smith, P. H. (**1985**). “Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers,” Hearing Research **18**(2), 159–168.

Robert, A. and Eriksson, J. L. (**1999**). “A composite model of the auditory periphery for simulating responses to complex sounds,” The Journal of the Acoustical Society of America **106**(4), 1852–1864.

Robles, L. and Ruggero, M. A. (**2001**). “Mechanics of the mammalian cochlea,” Physiological reviews **81**(3), 1305–1352.

Rose, J. E., Brugge, J. F., Anderson, D. J. and Hind, J. E. (**1967**). “Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey.,” Journal of neurophysiology **30**(4), 769–793.

Rose, J. E., Hind, J. E., Anderson, D. J. and Brugge, J. F. (**1971**). “Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey.,” Journal of Neurophysiology **34**(4), 685–699.

Rosen, S. (**1992**). “Temporal information in speech: acoustic, auditory and linguistic aspects,” Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **336**(1278), 367–373.

Rosowski, J. J. (**1996**). “Models of external- and middle-ear function,” *in* H. L Hawkins, T. A McMullen, A. N Popper and R. R Fay, eds, “Auditory Computation,” Springer New York New York, NY pp. 15–61.

Rothman, J. S. and Manis, P. B. (**2003**). “The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons.,” Journal of neurophysiology **89**(6), 3097–3113.

Ruggero, M. A. (**1992**). “Physiology and coding of sound in the auditory nerve,” *in* “The mammalian auditory pathway: Neurophysiology,” Springer pp. 34–93.

Saberi, K. (**1995**). “Some considerations on the use of adaptive methods for estimating interaural-delay thresholds,” The Journal of the Acoustical Society of America **98**(3), 1803–1806.

Sabin, A. T., Eddins, D. a. and Wright, B. a. (**2012**). “Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system.,” The Journal of neuroscience : the official journal of the Society for Neuroscience **32**(19), 6542–9.

Salminen, N. H., Jones, S. J., Christianson, G. B., Marquardt, T. and McAlpine, D. (**2018**). “A common periodic representation of interaural time differences in mammalian cortex,” NeuroImage **167**, 95–103.

Salvi, R., Perry, J., Hamernik, R. P. and Henderson, D. (**1982**). “Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma,” New perspectives on noise-induced hearing loss pp. 165–188.

Saoji, A. A., Litvak, L., Spahr, A. J. and Eddins, D. A. (**2009**). "Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners," Journal of the Acoustical Society of America **126**(3), 955–958.

Saremi, A., Beutelmann, R., Dietz, M., Ashida, G., Kretzberg, J. and Verhulst, S. (**2016**). "A comparative study of seven human cochlear filter models," The Journal of the Acoustical Society of America **140**(3), 1618–1634.

Sayers, B. M. and Cherry, E. C. (**1957**). "Mechanism of binaural fusion in the hearing of speech," The Journal of the Acoustical Society of America **29**(9), 973–987.

Sergent, M. (**2018**). "'Ear diagram,"'.
URL: *http://cssmith.co/ear-tympanic-membrane-diagram/*

Sergeyenko, Y., Lall, K., Liberman, M. C. and Kujawa, S. G. (**2013**). "Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline," Journal of Neuroscience **33**(34), 13686–13694.

Shackleton, T. M., Meddis, R. and Hewitt, M. J. (**1992**). "Across frequency integration in a model of lateralization," The Journal of the Acoustical Society of America **91**(4), 2276–2279.

Shamma, S. A., Chadwick, R. S., Wilbur, W. J., Morrish, K. A. and Rinzel, J. (**1986**). "A biophysical model of cochlear processing: Intensity dependence of pure tone responses," The Journal of the Acoustical Society of America **80**(1), 133–145.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**(5234), 303–304.

Shera, C. A., Guinan, J. J. and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proceedings of the National Academy of Sciences **99**(5), 3318–3323.

Sidwell, A. and Summerfield, Q. (**1986**). "The auditory representation of symmetrical cvc syllables," Speech Commun **5**(3), 283–297.

Siebert, W. M. (**1965**). "Some implications of the stochastic behavior of primary auditory neurons," Kybernetik **2**(5), 206–215.

Siebert, W. M. (**1968**). "Stimulus transformations in the peripheral auditory system," Recognizing patterns pp. 104–133.

Siebert, W. M. (**1970**). "Frequency discrimination in the auditory system: Place or periodicity mechanisms?," Proceedings of the IEEE **58**(5), 723–730.

Slaney, M. et al. (**1993**). "An efficient implementation of the patterson-holdsworth auditory filter bank," Apple Computer, Perception Group, Tech. Rep **35**(8).

Smith, Z. M., Delgutte, B. and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**(6876), 87–90.

Smoski, W. J. and Trahiotis, C. (**1986**). “Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss,” The Journal of the Acoustical Society of America **79**(5), 1541–1547.

Spencer, N. J., Hawley, M. L. and Colburn, H. S. (**2016**). “Relating interaural difference sensitivities for several parameters measured in normal-hearing and hearing-impaired listeners,” The Journal of the Acoustical Society of America **140**(3), 1783–1799.

Spoendlin, H. and Schrott, A. (**1988**). “The spiral ganglion and the innervation of the human organ of corti,” Acta oto-laryngologica **105**(5-6), 403–410.

Spoendlin, H. and Schrott, A. (**1989**). “Analysis of the human auditory nerve,” Hearing research **43**(1), 25–38.

Stecker, G. C. and Gallun, F. J. (**2012**). “Binaural hearing, sound localization, and spatial hearing,” *in* “Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing,” Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing Plural Publishing Ch. 14, 383 – 433.

Steele, C. R., Baker, G., Tolomeo, J. and Zetes, D. (**1993**). “Electro-mechanical models of the outer hair cell,” *in* D H., J. W Horst, P van Dijk and S. M van Netten, eds, “Biophysics of Hair-Cell Sensory Systems,”.

Steeneken, H. J. and Houtgast, T. (**1980**). “A physical method for measuring speech-transmission quality,” J. Acoust. Soc. Am. **67**(1), 318–326.

Stern, R. M. and Colburn, H. S. (**1978**). “Theory of binaural interaction based on auditory-nerve data. iv. a model for subjective lateral position,” The Journal of the Acoustical Society of America **64**(1), 127–140.

Stern, R. M. and Shear, G. D. (**1996**). “Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay,” The Journal of the Acoustical Society of America **100**(4), 2278–2288.

Stern, R. M. and Trahiotis, C. (**1992**). “The role of consistency of interaural timing over frequency in binaural lateralization,” Auditory physiology and perception pp. 547–554.

Stilp, C. E. and Kluender, K. R. (**2010**). “Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility,” Proceedings of the National Academy of Sciences **107**(27), 12387–12392.

Strange, W. and Bohn, O.-S. (**1998**). “Dynamic specification of coarticulated german vowels: Perceptual and acoustical studies,” J. Acoust. Soc. Am. **104**(1), 488–504.

Stüttgen, M. C., Schwarz, C. and Jäkel, F. (**2011**). “Mapping spikes to sensations,” Frontiers in Neuroscience **5**, 1–17.

Sumner, C. J., Lopez-Poveda, E. A., O’Mard, L. P. and Meddis, R. (**2002**). “A revised model of the inner-hair cell and auditory-nerve complex,” The Journal of the Acoustical Society of America **111**(5), 2178–2188.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P. and Meddis, R. (**2003**). "Adaptation in a revised inner-hair cell model," The Journal of the Acoustical Society of America **113**(2), 893–901.

Sun, Q., Gan, R., Chang, K.-H. and Dormer, K. (**2002**). "Computer-integrated finite element modeling of human middle ear," Biomechanics and modeling in mechanobiology **1**(2), 109–122.

Supin, A. Y., Popov, V. V., Milekhina, O. N. and Tarakanov, M. B. (**1997**). "Frequency-temporal resolution of hearing measured by rippled noise," Hearing Research **108**(1-2), 17–27.

Swaminathan, J. and Heinz, M. G. (**2012**). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," J. Neurosci. **32**(5), 1747–1756.

Takanen, M., Santala, O. and Pulkki, V. (**2014**). "Visualization of functional count-comparison-based binaural auditory model output," Hearing research **309**, 147–163.

Tan, Q. and Carney, L. H. (**2003**). "A phenomenological model for the responses of auditory-nerve fibers. ii. nonlinear tuning with a frequency glide," The Journal of the Acoustical Society of America **114**(4), 2007–2020.

van der Heijden, M., Lorteije, J. A., Plauška, A., Roberts, M. T., Golding, N. L. and Borst, J. G. G. (**2013**). "Directional hearing by linear summation of binaural inputs at the medial superior olive," Neuron **78**(5), 936–948.

Van Deun, L., Van Wieringen, A., Francart, T., Scherf, F., Dhooge, I. J., Deggouj, N., Desloovere, C., Van De Heyning, P. H., Offeciers, F. E., De Raeve, L. and Wouters, J. (**2009**). "Bilateral cochlear implants in children: Binaural unmasking," Audiology and Neurotology **14**(4), 240–247.

van Hoesel, R. J. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," The Journal of the Acoustical Society of America **121**(4), 2192–2206.

van Hoesel, R. J. and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects.," The Journal of the Acoustical Society of America **102**(1), 495–507.

van Hoesel, R. J., Jones, G. L. and Litovsky, R. Y. (**2009**). "Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation," Journal of the Association for Research in Otolaryngology **10**(4), 557.

van Hoesel, R. J. M. (**2012**). "Contrasting benefits from contralateral implants and hearing aids in cochlear implant users," **288**(1-2), 100–113.

van Hoesel, R. J. and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," The Journal of the Acoustical Society of America **113**(3), 1617–1630.

Van Wieringen, A. and Wouters, J. (**2008**). “List and lint: Sentences and numbers for quantifying speech understanding in severely impaired listeners for flanders and the netherlands,” Int. J. Audiol. **47**(6), 348–355.

Ventura, V. (**2010**). “Bootstrap tests of hypotheses,” *in* S Grün and S Rotter, eds, “Analysis of Parallel Spike Trains,” Vol. 7 Springer Ch. 18, 383–398.

Vercammen, C., Goossens, T., Undurraga, J., Wouters, J. and van Wieringen, A. (**2018**). “Electrophysiological and behavioral evidence of reduced binaural temporal processing in the aging and hearing impaired human auditory system,” Trends in hearing **22**, 1–12.

Verhulst, S., Altoè, A. and Vasilkov, V. (**2018**). “Computational modeling of the human auditory periphery: auditory-nerve responses, evoked potentials and hearing loss,” Hearing Research **360**, 55–75.

Verhulst, S., Dau, T. and Shera, C. A. (**2012**). “Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission,” The Journal of the Acoustical Society of America **132**(6), 3842–3848.

Viergever, M. A. (**1980**). Mechanics of the inner ear: a mathematical approach PhD thesis.

Walsh, T., Demkowicz, L. and Charles, R. (**2004**). “Boundary element modeling of the external human auditory system,” The Journal of the Acoustical Society of America **115**(3), 1033–1043.

Wang, D. and Brown, G. J. (**2005**). “Binaural sound localization,” *in* “Computational Auditory Scene Analysis,” John Wiley & Sons Ch. 5, 1–34.

Wang, K. and Shamma, S. (**1995**). “Spectral shape analysis in the central auditory system,” Speech and Audio Processing, IEEE Transactions on **3**(5), 382–395.

Wang, L. and Colburn, H. S. (**2012**). “A modeling study of the responses of the lateral superior olive to ipsilateral sinusoidally amplitude-modulated tones,” JARO - Journal of the Association for Research in Otolaryngology **13**(2), 249–267.

Wang, L., Devore, S., Delgutte, B. and Colburn, H. S. (**2014**). “Dual sensitivity of inferior colliculus neurons to itd in the envelopes of high-frequency sounds: experimental and modeling study,” Journal of neurophysiology **111**(1), 164–181.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. (**2004**). “Image quality assessment: from error visibility to structural similarity,” Image Processing, IEEE Transactions on **13**(4), 600–612.

Weisz, C., Glowatzki, E. and Fuchs, P. (**2009**). “The postsynaptic function of type ii cochlear afferents,” Nature **461**(7267), 1126.

Westerman, L. A. (**1985**). “Adaptation and recovery of auditory nerve responses,” Technical report Syracuse University.

Wightman, F. L. and Kistler, D. J. (**1992**). “The dominant role of low-frequency interaural time differences in sound localization.,” The Journal of the Acoustical Society of America **91**(3), 1648–61.

Williams, E. J. and Williams, E. (**1959**). *Regression analysis* Vol. 14 Wiley New York, NY, USA.

Winter, I. M., Robertson, D. and Yates, G. K. (**1990**). “Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres,” Hearing research **45**(3), 191–202.

Witten, I. and Frank, E. (**2011**). *Data Mining: Practical machine learning tools and techniques* third edn Morgan Kaufmann.

Won, J. H., Moon, I. J., Jin, S., Park, H., Woo, J., Cho, Y.-S., Chung, W.-H. and Hong, S. H. (**2015**). “Spectrotemporal modulation detection and speech perception by cochlear implant users,” PloS one **10**(10), e0140920.

Woo, J., Miller, C. A. and Abbas, P. J. (**2010**). “The dependence of auditory nerve rate adaptation on electric stimulus parameters, electrode position, and fiber diameter: a computer model study,” Journal of the Association for Research in Otolaryngology **11**(2), 283–296.

Wright, A., Davis, A., Bredberg, G., Ulehlova, L. and Spencer, H. (**1987**). “Hair cell distributions in the normal human cochlea.,” Acta oto-laryngologica. Supplementum **444**, 1–48.

Yin, T. C., Chan, J. C. and Irvine, D. R. (**1986**). “Effects of interaural time delays of noise stimuli on low-frequency cells in the cat’s inferior colliculus. i. responses to wideband noise,” Journal of Neurophysiology **55**(2), 280–300.

Young, E. D. (**2008**). “Neural representation of spectral and temporal information in speech,” Philosophical Transactions of the Royal Society B: Biological Sciences **363**(1493), 923–945.

Young, E. and Sachs, M. B. (**1973**). “Recovery from sound exposure in auditory-nerve fibers,” The Journal of the Acoustical Society of America **54**(6), 1535–1543.

Zhang, K. D. and Coate, T. M. (**2017**). “Recent advances in the development and function of type ii spiral ganglion neurons in the mammalian inner ear,” *in* “Seminars in cell & developmental biology,” Vol. 65 Elsevier pp. 80–87.

Zhang, T., Spahr, A. J., Dorman, M. F. and Saoji, A. (**2013**). “Relationship Between Auditory Function of Nonimplanted Ears and Bimodal Benefit,” Ear and hearing **34**(2), 133–41.

Zhang, X., Heinz, M. G., Bruce, I. C. and Carney, L. H. (**2001**). “A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression,” J. Acoust. Soc. Am. **109**(2), 648–670.

Zheng, Y., Escabí, M. and Litovsky, R. Y. (**2017**). “Spectro-temporal cues enhance modulation sensitivity in cochlear implant users,” Hearing Research **351**, 45–54.

Zilany, M. S. and Bruce, I. C. (**2006**). “Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery,” J. Acoust. Soc. Am. **120**(3), 1446–1466.

Zilany, M. S. and Bruce, I. C. (**2007**). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," *in* "Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on," IEEE Kohala Coast, HI, USA pp. 481–485.

Zilany, M. S., Bruce, I. C. and Carney, L. H. (**2014**). "Updated parameters and expanded simulation options for a model of the auditory periphery," J. Acoust. Soc. Am. **135**(1), 283–286.

Zilany, M. S., Bruce, I. C., Nelson, P. C. and Carney, L. H. (**2009**). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics," J. Acoust. Soc. Am. **126**(5), 2390–2412.

Zilany, M. S. and Carney, L. H. (**2010**). "Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics," J. Neurosci. **30**(31), 10380–10390.

Scientific Acknowledgements

Thanks to Jana Van Canneyt for helping with the translation of the Abstract from English to Dutch. For Ch. 2, we would like to thank Dr. Raphael Koning for providing the base scripts from which that work was developed as well as Prof. Dr. Torsten Dau and Dr. Alexandre Chabot-Leclerc for their help in the calibration of the sEPSM ideal observer. For Ch. 3, we would like to thank Benjamin Dieudonné for his help generating the stimuli. For Ch. 4, we would like to thank Prof. Dr. Ville Pulkki, Prof. Dr. Torsten Dau, Prof. Dr. Bastian Epp, and Dr. Suyash Joshi for stimulating idea exchanges. For Ch. 5, we would like to thank Dr. Christoph Scheidiger and Prof. Dr. Ian C. Bruce for the fruitful discussions.

We would also like to thank the editors and (anonymous) reviewers for their helpful comments and feedback. Finally, our most sincere thanks to all our participants for their time and dedicated effort.

Personal Contribution

In all studies described in this thesis, Arturo Moncada-Torres was responsible for the design, implementation, and evaluation of the computational models, as well as for the writing and revision of the corresponding manuscripts.

In Ch. 3, Arturo Moncada-Torres and Sara Magits declare equal contributions to the study and should be considered as joint first authors. In this chapter, Arturo Moncada-Torres was also responsible together with Sara Magits for the analysis of the behavioral data.

Conflict of Interest Statement

The (co-)authors of all chapters declare that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this thesis.

Arturo Moncada-Torres received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. 317521 (ICanHear) and from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 637424 (ERC Starting Grant to Tom Francart).

Co-author of Ch. 3 Sara Magits received funding from a TBM-FWO grant from the Research Foundation-Flanders (grant no. T002216N), Flanders Innovation & Entrepreneurship (formerly IWT) for project no. 50432, and Cochlear Ltd.

Co-author of Ch. 4 Dr. Andreas Prokopiou received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. 317521 (ICanHear) and from the Flanders Innovation & Entrepreneurship (formerly IWT) for project no. 150432.

Co-author of Ch. 5 Dr. Suyash N. Joshi received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no. 317521 (ICanHear).

The computational resources and services used in Ch. 4 and 5 were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government Department EWI.

Curriculum Vitae

Arturo Moncada-Torres was born on February 15, 1988 in Mexico City, Mexico. In 2012, he received his Bachelor of Science in Biomedical Engineering (with focus in Instrumentation) from the Universidad Iberoamericana (Mexico City, Mexico). In 2014, he received his Master of Science in Biomedical Engineering (with focus in Bioimaging) from the ETH Zürich (Zürich, Switzerland).

He joined the group Experimental Otorhinolaryngology (ExpORL) at the Department of Neurosciences at the University of Leuven (Leuven, Belgium) on July 2014 as a Doctoral Researcher under the supervision of Prof. Dr. Tom Francart . He was part of the Marie Cure Initial Training Network *Improved Communication through Applied Hearing Research* (ICanHear). Currently, he is a Clinical Data Scientist in the Integral Kankercentrum Nederland (IKNL) in Eindhoven, The Netherlands. His personal interests include latin, swing, and ballroom dancing, cooking (and eating), and volleyball.

List of Publications

Related to this Thesis

Peer-reviewed Journal Papers

- Moncada-Torres, A., van Wieringen, A., Bruce, I.C., Wouters, J., and Francart, T. (**2017**). “Predicting phoneme and word recognition in noise using a computational model of the auditory periphery,” The Journal of the Acoustical Society of America, **141**(1):300–312.
- Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. (**2017**). “Functional modelling of interaural time difference discrimination in acoustical and electrical hearing,” Journal of Neural Engineering, **14**(4):1-21.
- Moncada-Torres, A., Joshi, S. N., Prokopiou, A., Wouters, J., Epp, B., and Francart, T. (**2018**). “A framework for computational modelling of interaural time difference discrimination of normal and hearing-impaired listeners.” The Journal of the Acoustical Society of America, **144**(2):940-954.
- Moncada-Torres, A., Sara Magits, Van Deun, L., Wouters, J., and Francart, T. (**2018**). “The effect of presentation level on spectrotemporal modulation detection.” Hearing Research. *In press.*

Peer-/Panel-reviewed Posters/Talks

- Moncada-Torres, A., Joshi, S. N., Epp, B., and Francart, T. “Decision device comparison for model-based analysis of ITD perception in normal hearing listeners.” In *ARCHES 2016 + ICanHearConference*, Zurich, Switzerland, November 2016.
- Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. “Functional modelling of just noticeable differences in interaural time difference discrimination for bimodal and bilateral cochlear implant users.” In *ARCHES + ICanHearConference*, Zurich, Switzerland, November 2016.
- Moncada-Torres, A., Joshi, S. N., Epp, B., and Francart, T. “Model-based analysis of ITD perception in normal & hearing impaired listeners.” In *International Hearing Aid Reasearch Conference (IHCON)*, Granlibakken, CA, USA, August 2016.
- Prokopiou, A., Moncada-Torres, A., Wouters, J., and Francart, T. “Functional modelling of just noticeable differences in interaural time difference discrimination for bimodal and bilateral cochlear implant stimulation.” In *Workshop on Auditory Neuroscience, Cognition, and Modelling*, London, UK, February 2016.
- Moncada-Torres, A., Epp, B., Francart, T., and Joshi, S. N. “Model-based analysis of interaural time differences in auditory nerve responses.” In *ISAAR 2015*, Nyborg, Denmark, August 2015.
- Moncada-Torres, A., Koning, R., Wouters, J., and Francart, T. “Predicting phoneme and word recognition using a computational model of the normal auditory periphery.” In *The Auditory Modelling Workshop*, Oldenburg, Germany, June 2015.
- Moncada-Torres, A., Koning, R., Wouters, J., and Francart, T. “Predicting phoneme and word recognition using a computational model in normal-hearing listeners.” In *ARCHES 2014*, Oldenburg, Germany, November 2014.

Unrelated to this Thesis

Peer-reviewed Journal Papers

- Van Hirtum, T., Moncada-Torres, A., Ghesquière, P., and Wouters, J. (**2018**). "Speech envelope enhancement instantaneously restores atypical speech perception in dyslexia." *Under review.*

Peer-/Panel-reviewed Posters/Talks

- Van Hirtum, T., Moncada-Torres, A., Ghesquière, P., and Wouters, J. "Exploring Speech Envelope Enhancement in Adults with Dyslexia." In *SPIN 2018*, Glasgow, UK, January 2018.
- Van Hirtum, T., Moncada-Torres, A., Ghesquière, P., and Wouters, J. "The potential of speech envelope enhancement for auditory intervention in dyslexia." In *International Hearing Aid Reasearch Conference (IHCON)*, Granlibakken, CA, USA, August 2018.

University of Leuven
Faculty of Medicine
Department of Neurosciences
ExpORL
Herestraat 49, Bus 721
B-3000 Leuven