Characterization and Classification of the Frequency Following Response to Vowels at Different Sound Levels in Normal Hearing Adults

by

Brian Heffernan

A thesis submitted to the Faculty of Graduate and Postgraduate Studies
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Rehabilitation Sciences

Faculty of Health Sciences
School of Rehabilitation Sciences
University of Ottawa
Ottawa, Canada
January 2019

Abstract

This work seeks to more fully characterize how the representation of English vowels changes with increasing sound level in the frequency following response (FFR) of normal-hearing adult subjects. It further seeks to help inform the design of brain-computer interfaces (BCI) that exploit the FFR for hearing aid (HA) applications. The results of three studies are presented, followed by a theoretical examination of the potential BCI space as it relates to HA design.

The first study examines how the representation of a long vowel changes with level in normal hearing subjects. The second study examines how the representation of four short vowels changes with level in normal hearing subjects. The third study utilizes machine learning techniques to automatically classify the FFRs captured in the second study. Based in-part on the findings from these three studies, potential avenues to pursue with respect to the utilization of the FFR in the automated fitting of HAs are proposed.

The results of the first two studies suggest that the FFR to vowel stimuli presented at levels in the typical speech range provide robust and differentiable representations of both envelope and temporal fine structure cues present in the stimuli in both the time and frequency domains. The envelope FFR at the fundamental frequency (F0) was generally not monotonic-increasing with level increases. The growth of the harmonics of F0 in the envelope FFR were consistent indicators of level-related effects, and the harmonics related to the first and second formants were also consistent indicators of level effects.

The third study indicates that common machine-learning classification algorithms are able to exploit features extracted from the FFR, both in the time and frequency domains, in order to accurately predict both vowel and level classes among responses. This has positive implications for future work regarding BCI-based approaches to HA fitting, where controlling for clarity and loudness are important considerations.

Dedication

To my mom for all of her support and motivation. To Sarah for all of her love throughout our great adventure together. And in loving memory of my dad, who imparted to me the grit and determination required to make it through the gauntlet.

Acknowledgements

I would like to thank my supervisors, Drs. Hilmi Dajani and Christian Giguère for their support, direction, and great patience without which this work would not have been possible. Je voudrais aussi dire merci aux professeurs Amineh Korivand, Josée Lagacé, et André Marcoux pour leurs temps et leurs commentaires. Finally, I would like to thank all of my colleagues at RGN over the years for their support and friendship, and all of the subjects who participated in my studies for their time.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABR	auditory brainstem response.
AC	auditory cortex.
AEP	auditory evoked potential.
am	amplitude modulation.
AN	auditory nerve.
ASSR	auditory steady state response.
BCI	brain-computer interface.
BCI-HA	brain-computer interfaced hearing-aid.
cABR	complex auditory brainstem response.
CAEP	cortical auditory evoked potential.
CN	cochlear nucleus.
CV	cross-validation.
dB	decibel.
dB nHL	decibels - normalized hearing level.
dB SL	decibels - sensation level.
dB SPL	decibels - sound pressure level.
dBA	decibels - A-weighted.
DFT	discrete Fourier transform.
DT	decision tree.
EEG	electroencephalography.
eFFR	envelope frequency following response.
EFR	envelope following response.
ENV	envelope.
F0	fundamental frequency.
F1	first formant.
F2	second formant.

FFR	frequency following response.
fMRI	functional magnetic resonance imaging.
GM	grand mean.
H_i	The ith harmonic of F0.
HA	hearing aid.
HI	hearing impaired.
HL	hearing loss.
IC	inferior colliculus.
IHC	inner hair cell.
ISI	inter-stimulus interval.
KNN	k-nearest neighbors.
LDA	linear discriminant analysis.
LL	lateral lemniscus.
LLAEP	long latency auditory evoked potential.
LOO	leave-one-out.
LOSO	leave-one-subject-out.
LR	logistic regression.
MEG	magnetoencephalography.
MGB	medial geniculate body.
ML	machine learning.
MLR	middle latency response.
RF	random forest.
RMANOVA	repeated-measures analysis of variance.
RMS	root-mean-square.
sFFR	spectral frequency following response.

SFHA self-fitting hearing aid.

SIN speech-in-noise.

SNHL sensorineural hearing loss.

SNR signal-to-noise ratio.

SVM support vector machine.

TFS temporal fine structure.

Chapter 1
Introduction

1.1 Context

Hearing impairment is a serious global health problem from which more than 360 million people currently suffer (Olusanya, Neumann, & Saunders, 2014). This represents roughly 5.3% of the world's population, and makes hearing impairment the 3rd most common cause of years lived with disability. Since the overwhelming majority of the hearing impaired (HI) population live in developing countries, one might expect to see a decline in the prevalence of hearing impairment as we push towards a more developed, global economy. Unfortunately, the evidence is quite to the contrary, and due in part to an aging population and to a variety of techno-industrial trends, it is projected that the size of the impaired population will continue to grow rapidly in the decades to come (McPherson, 2011). In Canada and the United States alone, it is estimated that 15-20% of adults have some trouble hearing (Feder, Michaud, Ramage-Morin, McNamee, & Beauregard, 2015), and a recent study by the World Health Organization estimates that 1.1 billion teenagers and young adults are at increased risk of hearing loss (HL) due to their exposure to loud environments such as nightclubs and to the ubiquity of smartphone usage at unsafe sound levels (World Health Organization et al., 2015).

Despite its prevalence, hearing loss often occurs very gradually as we age, and many people are simply unaware of their impairment (Amieva et al., 2015; Lin, 2012). Even among those who are aware of their hearing loss, very few seek any form of treatment, as estimates show that roughly one fifth of older adults with a hearing loss use a hearing aid (Lin, Thorpe, Gordon-Salant, & Ferrucci, 2011). This is to the detriment of hearing impaired individuals and their families as well as society as a whole, as a variety of negative impacts – both cognitive and physical – have been associated with such impairment (Kamil & Lin, 2015; Smith, 2008). These include increased risk for dementia, depression, cognitive decline and social withdrawal, as well as an increased likelihood of hospitalization and mortality in general (Amieva et al., 2015; Contrera et al., 2017; Genther et al., 2015; Genther, Frick, Chen, Betz, & Lin, 2013; Lin & Albert, 2014; Lin et al., 2013; Mick, Kawachi, & Lin, 2014). Furthermore, persons with hearing loss are at an increased risk of low educational attainment and poverty across the globe (Emmett & Francis, 2015).

The most common treatment of hearing impairment and just as commonly the most suitable is the prescription of a well-known assistive technology – the hearing aid (HA) (Bainbridge & Ramachandran, 2014). By way of an audiological assessment to determine the nature of the impairment, an amplification profile is prescribed based on the selected fitting protocol, the specifics of the particular hearing aid(s) selected to be worn, and often some measure of feedback from the wearer (Beck & Nilsson, 2013). Although significant strides have been made towards improving HA technology and the satisfaction of the user experience with respect to the use of these devices over the past decades, a significant degree of dissatisfaction still exists across multiple dimensions of the typical user experience (Kochkin, 2010), which begs the question of whether or not there are perhaps new or novel approaches towards the development or the fitting of the technology that ought to be explored.

Among these approaches, the self-fitting hearing aid (SFHA) seems to be particularly promising. Not only do SFHAs appear capable of greatly reducing the barriers to access across a variety of demographics due to the diminished need for professional audiological services and the cost reductions associated with the devices themselves, their utilization also appears to result in positive impacts owing to increased user-engagement and the sense of self-care that accompanies it (Keidser & Convery, 2016).

Another approach has taken the objective of creating improved assessment and treatment approaches by making use of neurological correlates of auditory processing capabilities (or the lack thereof), and a few lines of attack have been put to the test in recent years. One such line involves the use of the click-evoked auditory brainstem response – a stereotyped response that is a hallmark of early-stage neural auditory processing, and a response that is well characterized among healthy adults. In its typical usage, this brainstem response is only of practical utility when the determination of behavioural thresholds is otherwise difficult to obtain (e.g. in persons with dementia) or is entirely unobtainable (e.g. in infants) (D. Stapells, Gravel, & Martin, 1996; Zaitoun, Purcell, & Cumming, 2015).

Indeed, the vast majority of these neural correlates have been primarily investigated for their utility as proxies for threshold determination, and relatively little has been done to explore their potential role(s) in the fitting process beyond this. This is also true of the cortical auditory evoked potential (CAEP) (Carter, Dillon, Seymour, Seeto, & Van Dun, 2013; Chun, Billings, Miller, & Tremblay, 2016), the auditory steady state response (ASSR) (Ahn, Lee, Kim, Yoon, & Chung, 2007; Hatzopoulos et al., 2009; Korczak, Smart, Delgado, M Strobel, & Bradford, 2012), and of the tonal frequency following response (FFR) as well, albeit to a much lesser extent (Lewis, Kopun, Neely, Schmid, & Gorga, 2015; Mobley & Gibson, 2000).

Among the research that has explored the utility of neurophysiological responses in the

fitting process, most have focused on evaluating the utility of CAEPs (C. J. Billings, Tremblay, & Miller, 2011) and of the ASSR (Zenker Castro & Barajas de Prat, 2008). Despite these efforts, how HA factors contribute to the aided CAEP is still not well understood, and although the ASSR shows promise not just in its utility to estimate thresholds in the absence of behavioural responses but also in the setting of HA features such as the gain and the compression ratio, and in estimating loudness sensation in aided conditions, more empirical demonstrations of their effectiveness and accuracy is required, especially in light of its unacceptably large estimation error in the proposed areas of clinical utility.

In keeping with audiological tradition, the tools built around the above-mentioned neurological correlates have made use of stimuli that are very 'artificial' in that they are not only synthesized, but that they do not usually occur in nature and are not always the most appropriate proxies for other behaviourally-relevant, frequently occurring sounds either (Purves et al., 2012). Since two of the largest problems identified by hearing aid users the world over include issues related to controlling loudness and difficulty comprehending speech-in-noise (SIN), and owing to our understanding that these are both complex, non-linear processes that cannot be sufficiently predicted with simple stimuli, there has been a recent push towards making use of more ecologically valid stimuli such as speech sounds themselves (Beck & Nilsson, 2013).

Although there is no generally accepted explanation of how the subjective percept of loudness is encoded based on the sound level of a stimulus, some recent and very high-quality fMRI and MEG studies seem to indicate that a transformation from level-coding to loudness-coding occurs somewhere between the rostral brainstem (e.g. the inferior colliculus) and the auditory cortex (Behler & Uppenkamp, 2016a, 2016b; Hart, Hall, & Palmer, 2003; Langers, van Dijk, Schoenmaker, & Backes, 2007; Röhl & Uppenkamp, 2012; Thwaites, Glasberg, Nimmo-Smith, Marslen-Wilson, & Moore, 2016; Uppenkamp & Röhl, 2014) (see Appendix A.4.5). This is in keeping with our general notion of the organization of the brain with respect to the emergence of perceptual correlates in the cortical regions more broadly across all manner of sensory inputs (de Lafuente & Romo, 2005; Vázquez, Salinas, & Romo, 2013; X. Wang, Lu, Bendor, & Bartlett, 2008).

1.2 Motivation and aims

Given the above-mentioned coding transformation from level to loudness, it is natural to wonder whether or not one is able to characterize the normal encoding of level in the brainstem, and if so, whether or not this information could then be used as a target or otherwise exploited somehow in order to (automatically) fit HAs. And if so, would this then result

in an 'optimal' or 'more natural' encoding of loudness for HA users, and would this then translate into better speech comprehension in both quiet and in noise, all else being equal?

Clearly, the first of these questions must be addressed before it is possible to address the latter thus posed. Besides the fact that a solid understanding of normal responses is necessary to characterize abnormal or pathological responses and their degree of severity, it stands to reason that if there does not appear to be sufficient information in the responses of normal-hearing subjects for useful classification schemes, then it is unlikely that there would be sufficient information in the responses of hearing-impaired subjects for use in informing fitting schemes, as they generally have less specific representations of both frequency and level by comparison (by definition), and this cannot be fully remedied through the use of hearing aids. Of course, if the former bears out, then it will only serve to strengthen the call to gain deeper insights into the responses of hearing-impaired subjects in the future, both in aided and unaided conditions and across a variety of hearing aids and settings, as such an understanding will help to further inform and constrain automatic fittings.

So given that the electrophysiological activity of the brainstem seems a reasonable candidate target for gleaning information relevant to the encoding of level and loudness, and so therefore to the programming of hearing aids in theory, and that there is no more ecologically valid set of stimuli than speech sounds themselves, it is thus that I have undertaken research towards characterizing and automatically classifying the speech-evoked frequency following response of normal-hearing adults as a first and necessary step. Specifically, this work aims to provide a much more thorough characterization of both the time and frequency domain representations of both the spectral and envelope frequency following responses to speech at various sound levels.

Removed from its pragmatic considerations, the question of how level and loudness of speech is coded is itself of great importance from a pure auditory neuroscience research perspective. Although we have learned a lot about a variety of the electrophysiological responses of the auditory system over the past few decades, we as researchers need to possess a very in-depth understanding of the neural responses typically elicited by the stimuli that naturally occur in our environments in order to build the bridge towards a comprehensive neurophenomenology. Achieving such a view would enable us to infer the qualities of subjective perceptual experience with a good degree of accuracy and precision based upon neuronal responses alone, and would bring us one step closer to a comprehensive, 'full' theory of audition (see Fig. 1.1).

Although some work has been done to assess the classification accuracy of English vowel encoding in the FFR using some well-known ML algorithms and related modeling techniques (Jafarpisheh et al., 2016; Sadeghian, Dajani, & Chan, 2011, 2015; Won et al., 2016; Yi,

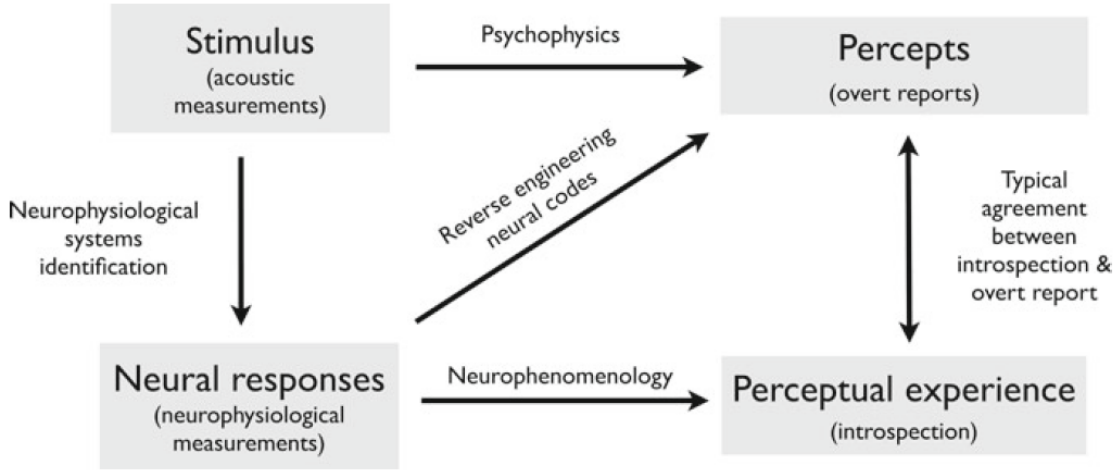

Figure 1.1: A full theory of audition must necessarily explain the relations between sounds, the neural responses that they induce, the functions of the auditory system, and the subject's auditory experience. The various disciplines involved in the study of these relations, both within and between components are noted in the figure. Reprinted with permission from "Toward a Theory of Information Processing in Auditory Cortex", by P. Cariani and C. Micheyl. In The Human Auditory Cortex (p. 354), by D. Poeppel et al. (Eds.), 2012, New York, NY: Springer New York. Copyright 2012 by Springer Science+Business Media LLC.

Xie, Reetzke, Dimakis, & Chandrasekaran, 2017), to my awareness, no study to date has looked at the classification of level encoding in the speech-evoked FFR, nor has any thorough assessment of the utility of the FFR in an BCI-HA construction ever been produced. It is thus that further to the characterization and classification of level effects in the speech-evoked FFR, this work aims to explore the feasibility – from a purely theoretical level, and based on the results from said FFR investigations – of using this very signal in the construction of a brain-computer interface (BCI) that would serve to automatically adjust the hearing aid profile, and that might therefore be able to account for changes owing to neural plasticity in addition to potentially accommodating changes due to user preference(s).

Here, not only is the FFR signal itself considered, but additional considerations related to advances in sensor and battery technology, hearing aid design, machine learning techniques for classification, etc. are provided as well. This is done in order to present a solid overview of the space in terms of its trends and the various bottlenecks and limitations that will likely need to be addressed prior to the realization of any consumer-ready technology that is ripe for adoption. A schematic of such a system - referred to throughout as a BCI-HA - is shown in Fig. 1.2.

To successfully create such a system – if it is indeed feasible – would likely result not only in increasingly better fittings over time, ultimately converging towards a near-optimal solution that modifies itself in response to changes in the user's hearing and (plastic) neural processing abilities, but it would likely also reduce or entirely remove many of the barriers to HA acquisition and utilization as well, especially in developing countries where audiological

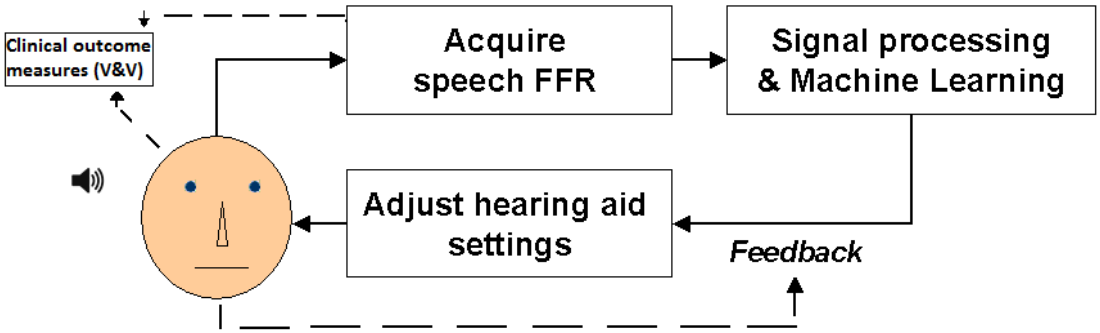

Figure 1.2: Schematic of a BCI-HA. Solid lines depict fundamental components; dashed lines represent potential human-based feedbacks (either clinician or user driven). User-based feedback could be realized via smartphone, for instance. Classical audiological outcome measures (i.e. verification and validation - or V&V) would allow the clinician to initialize the system appropriately, and could be used to interactively guide the system towards an optimal fitting.

services are sparse and cheap solutions are most desperately required. Indeed, there is good reason to believe that any such BCI-HAs would confer the same benefits to their end-users as those that appear to be experienced by the users of the SFHAs that have emerged in the last couple of years (although there is admittedly a need for many more quality studies of their performance and of user-satisfaction here).

This self-fitting approach to the next generation of HA technology, regardless of how it is arrived at, is all but guaranteed to persist, as is evidenced by a report from the American President's Council of Advisors on Science and Technology that was published in October 2016 (Keidser & Convery, 2016). The report states unequivocally that the direct-to-consumer sale of hearing aids can lower the cost of devices and ought to be approved, and that the cost of hearing services should be made more affordable and transparent by introducing an unbundled pricing structure for hearing health care. Given that this is a well-identified issue within one of the world's most developed countries, and that some of the self-fitted models appear to be comparable to other HAs in their performance once fit, it is obvious that the same approach would be extremely beneficial in less wealthy areas across the globe as well, and that the successful addition of a BCI component might further improve health outcomes and market penetration.

1.3 Research goal and research questions

In brief, my research goal was:

> *To more fully characterize how the representation of English vowels changes with increasing sound level in the frequency following response of normal-hearing adult subjects, with the secondary aim of helping to inform the design of brain-computer interfaces that exploit said response.*

Research questions

In line with this goal, a research plan was developed in order to answer the four following questions:

1. How does the representation of a long vowel change with level in the FFR of normal hearing subjects? [Full study]

2. How does the representation of four short vowels change with level in the FFR of normal hearing subjects? Do these representations differ between sexes? Are they stable across test and retest conditions? [Full study]

3. Can machine learning methods discriminate between FFR responses to different vowels AND/OR levels with good accuracy? [Full study]

4. Based in-part on the findings from the three previous studies, what are some potential avenues to pursue with respect to utilizing the FFR to inform the automated fitting of hearing aids? Which might be the best to pursue and why? [Conceptual design overview]

1.4 Contributions

The major findings pursuant to the work carried out in order to answer these four research questions are as follows:

1. The FFRs to vowel stimuli presented at levels typical of the speech range provide robust and differentiable representations of both envelope and temporal fine structure cues present in the stimuli via the envelope FFR and spectral FFR in both the time and frequency domains.

2. The growth of envelope FFR representations of the fundamental frequency (F0) is generally not monotonic-increasing with level in normal hearing subjects in quiet.

3. The growth of the second through to the sixth harmonics of F0 (denoted as H2-H6, where H_i represents the i^{th} harmonic of F0) in the envelope FFR were the most consistent statistical indicators of level-related effects. The growth of the spectral FFR harmonics related to the first and second formants (F1 and F2) were also consistent indicators of level effects.

4. Although small effects of sex are apparent in the amplitude spectra and major waveform peaks of speech-evoked FFRs, very few sex-level interactions appear to exist among normal hearing adults.

5. A variety of common classification algorithms are capable of achieving meaningful classification accuracies on both vowel, level, and vowel-level classification tasks when trained and tested on features extracted from the FFRs of normal hearing subjects in quiet.

6. On balance, the envelope FFR appears to provide more robust information related to presentation level classification accuracy than the spectral FFR. The spectral FFR appears to provide more robust information than the envelope FFR for the accurate classification of vowels.

7. Converging evidence based on the recent literature and the novel results above suggest that a variety of algorithmic approaches to FFR-based BCI-HA fitting are worthy of exploration.

1.5 Structure

This thesis is a monograph comprised of seven chapters, and its structure flows naturally as a consequence of the research questions outlined above. The remaining six chapters are organized as follows:

1. Chapter 2 provides background information pertaining to the FFR and the literature relevant to the studies addressing the first two research questions.

2. Chapter 3 provides background information pertaining to machine learning and brain-computer interfacing and the literature relevant to the third and fourth research questions.

3. Chapter 4 details the methods, results, and analyses of the two FFR studies.

4. Chapter 5 details the methods, results, and analyses of the machine learning study.

5. Chapter 6 summarizes a variety of potential BCI approaches to hearing aid fitting that seek to exploit the FFR. They are informed by the outcomes of the previously mentioned studies.

6. Chapter 7 provides a general conclusion based on the results of the work described herein, and provides suggestions for future work.

Chapter 2
Background: the frequency following response

For sake of brevity, a general systems-level overview of the auditory system – with a particular focus on speech and sound level processing – is provided in Appendix A instead of the main body of the thesis. This chapter commences with an overview of auditory evoked potentials (AEPs) and of the methodological considerations that pertain to the FFR prior to undertaking a focused literature review and a brief review of sex differences in the FFR. Finally, an analysis of the gaps present in the literature is provided.

2.1 Overview of human auditory evoked potentials

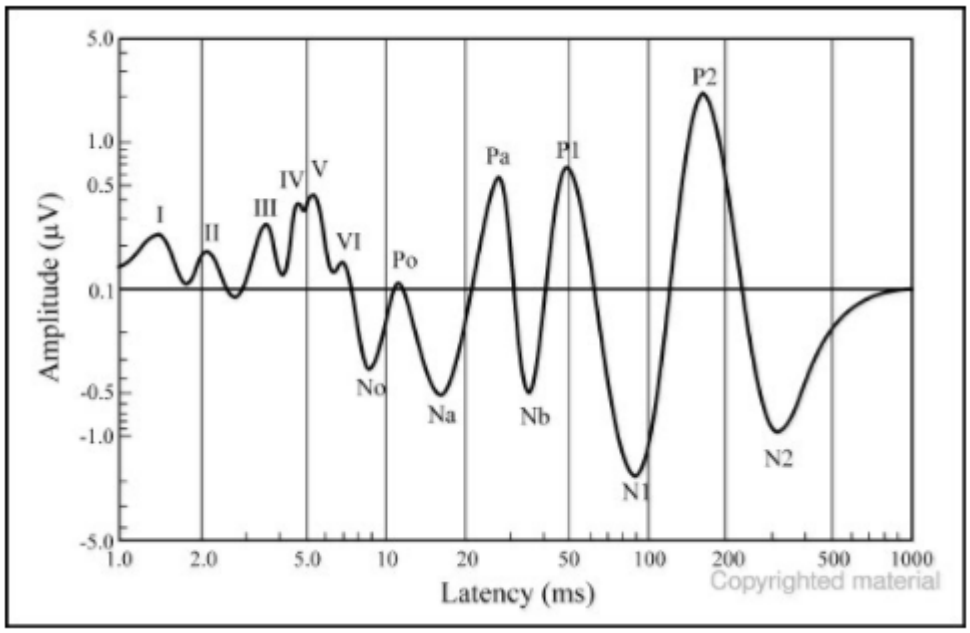

Figure 2.1: Auditory evoked potentials. Transient auditory evoked potentials to a click presented at 60 dB nHL and plotted on logarithmic time axis for visual clarity. Waves labelled as I-VI are the early response (i.e. the ABR); waves from No-Nb are middle-latency responses (MLR); waves P1-N2 constitute the late response (LLAEP). See Table 2.1 for a more thorough breakdown of both the transient responses shown here as well as the steady state and sustained responses, both further divided by latency. Reprinted with permission from Human Auditory Evoked Potentials (pg. 3), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

AEPs are a subclass of event-related potentials - neural responses that are time-locked to some event (e.g. a sensory stimulus or a mental event such as the recognition of a target stimulus) or the omission of a stimulus (Picton, 2010). For AEPs, the event is a sound,

and the elicited response to the sound manifests as very small electrical voltage potentials originating from the brain that are typically recorded from the scalp, although they can be recorded using invasive techniques as well. The selected stimuli vary, but are most typically clicks, tones, simple speech sounds, or noise.

There are a variety of AEPs that have been defined and characterized over the past several decades, the most common of which are in response to brief stimuli, such as clicks, shown in Fig. 2.1 along with their respective characteristic amplitude and latency. AEPs to brief stimuli are often divided up into three temporal groupings – early, middle, and late latency responses, as indicated in Fig. 2.1. This is roughly reflective of the relative layout of the neural architecture involved in processing sound as a stimulus ascends the auditory pathway (i.e. from the periphery and brainstem to the midbrain and finally to the cortex). In addition to classification via latency considerations, AEPs can be classified as either being transient (occurring following a change in stimulus such as an onset or offset), or sustained (evoked via continuous stimulation). This more complete classification scheme is summarized in Table 2.1, along with a brief description and examples of the real-world applicability of each AEP (see Burkard, Eggermont, and Don (2007) and Picton (2010) for comprehensive introductory texts on human AEPs).

It should come as no surprise that a broad subset of AEPs have been investigated both for purely scientific purposes as well as with respect to their potential clinical and audiological utility. In the latter setting, AEPs have typically been assessed with respect to their ability to somehow facilitate or ameliorate the fitting of HAs, especially in difficult to test populations such as infants or persons with cognitive impairments (Aiken & Picton, 2008; C. J. Billings, 2013). As part of the general characterization of AEPs, many studies have looked at the effects of stimulus level on amplitude and latency. While the literature suggests that a variety of relationships exist, there is a trend across all of the AEPs indicating that as level increases, latency generally decreases, while amplitude generally increases (see Appendix B for plots of said relationships across a variety of well-studied AEPs). Although this relationship is addressed in the literature review below as it pertains to the FFR, it is worth noting that this same typical pattern holds true for the click-evoked auditory brainstem response (ABR), and the latency and amplitude of waves I-V of the ABR can often be analyzed across these two very dimensions in order to screen for pathological processing in the central auditory system (Burkard et al., 2007; Picton, 2010).

Among the studies that have looked at AEPs as viable candidates for utilization in fitting processes in some form or another, some have borne fruit. Of note, the ABR can be used as a proxy for determining threshold (Mobley & Gibson, 2000; D. Stapells et al., 1996), as can be the auditory steady-state response (ASSR) (Hatzopoulos et al., 2009; Korczak et al.,

Table 2.1: Classification of auditory evoked potentials by latency and type. Transient responses occur following a change in the stimulus, such as its onset or offset. Sustained responses are evoked throughout the entirety of a stimulus. Steady-state responses are evoked by a regularly changing stimulus, and change regularly as well. Reprinted with permission from Human Auditory Evoked Potentials (pg. 5), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

Latency	Transient	Steady-State	Sustained
First (0-5 ms)	Cochlear Nerve Compound Action Potential	Cochlear Microphonic	Summating Potential
Fast (1-15 ms)	Auditory Brainstem Response	Frequency Following Response; Fast (>70 Hz) Auditory Steady-State Response	Pedestal of Frequency-Following Response
Middle (10-50 ms)	Middle-Latency Response	40-Hz Potential	
Slow (30-500 ms)	Vertex Potential	Slow (<30 Hz) Auditory Steady-State Response	Cortical Sustained Potential
Late (200-1000 ms)	Mismatch Negativity; Processing Negativity; Late Positive Waves		Contingent Negative Variation

2012). There have also been a handful of attempts to use the cortical AEPs to characterize threshold and normal processing. All of these studies, as well as the work that has been done to more broadly characterize the evoked potentials of the auditory system provide important context for the work presented herein and for all future work in the field.

2.2 The frequency following response

The nature of the FFR remains a contentious subject, despite having first been recorded over 40 years ago. Many of the points of contention appear to revolve around terminological issues, which are at least partly attributable to a lack of clear and standardized definitions within the discipline and to the inevitable conflation of terms that results. Indeed, a quick scan of the literature reveals the ongoing battle for terminological clarity and rigor, with several variations on the FFR theme and its related subcomponents appearing in works

from the last decade alone (see Kraus, Anderson, White-Schwoch, Fay, & Popper, 2017, for an overview). Further adding to the problem is the fact that our understanding of the neural sources (or generators) that contribute to the FFR has been challenged several times over the past few decades due to advances in the methods and techniques of recording and analyzing neurophysiological activity (see Bidelman, 2015, 2018; Chandrasekaran & Kraus, 2010; Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016; Coffey, Musacchia, & Zatorre, 2017; Tichko & Skoe, 2017).

Still to this day there are those who posit that the FFR might merely reflect a succession of onset responses, or in other words, that it is effectively little more than a summated (sustained) version of the ABR, and continued modeling is being done to assess this assertion (Bidelman, 2015). This notion is indicative of the long-held belief that the FFR is purely of brainstem origin, as the click-ABR as typically used in clinic is itself known to be devoid of any mid-brain or cortical contributions due to the brevity of the recording window. Any comprehensive search of the literature therefore demands the addition of terms such as the 'speech ABR', which would likely yield as many results as an FFR search itself in many subtopics of interest.

Even the authoritative tutorial published on the subject of FFR recording and analysis in 2010 is itself titled 'Auditory brain stem response to complex sounds: a tutorial' (Skoe & Kraus, 2010). This same paper introduces yet another acronym – the complex auditory brainstem response (cABR) – in order to distinguish the ABR to complex sounds from the traditional click-evoked ABR, and it partitions the cABR into two subcomponents: the spectral response and the envelope response. Unfortunately, no convenient terminology is introduced here in order to facilitate the easy differentiation between the two responses; however, it is common in the literature for the former to be referred to simply as the FFR, whereas the latter is referred to as the EFR (or the envelope following response) (Bharadwaj & Shinn-Cunningham, 2014; Dolphin & Mountain, 1992)! Further adding to the confusion, the envelope response has been referred to as the amplitude-modulated following response (Bharadwaj & Shinn-Cunningham, 2014; Kuwada et al., 2002). Bharadwaj and Shinn-Cunningham (2014), in an effort to cleanly differentiate between the tone-evoked (narrowband) FFR as well as responses containing cortical contributions, defined the auditory subcortical steady-state response (SSSRs). Their intention here is to further facilitate the differentiation of this response from responses to amplitude-modulated sounds originating from both the subcortical and cortical portions of the auditory pathway, which are collectively referred to as the auditory steady-state responses (ASSR).

Fortunately, studies from 2016 appear to point towards a possible resolution to all of the aforementioned issues related to both the terminology and of the open question of source

contributions to the FFR. Interesting work using electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) by Coffey, Herholz, et al. (2016) indicates that the FFR represents an integrated response from across the entirety of the auditory system, and that any speech token of sufficiently long duration (i.e. $>$ 80 ms) is all but guaranteed to elicit some cortical contributions. The implications here are not insubstantial, as they will inevitably impact considerations with respect to experimental setup, including recording montages (e.g. in order to highlight or diminish this right-hemispheric cortical contribution), stimulus creation (e.g. in order to reduce or enhance the likelihood of synchronous cortical contributions), and they demand the re-examination of a great number of findings that analyzed their results on the assumption of exclusive subcortical contributions. This is especially the case when effects of attention or other top-down cognitive mechanisms of cortical origin are known to be at play, and where timing and latency peaks may be influenced by constructive or destructive interference arising from the cortical components (Tichko & Skoe, 2017). Although these results are likely to result in some reframing of perspectives on the FFR, it bears mentioning here that a very recent study by Bidelman (2018) suggests that even though some cortical contributions are captured via the FFR, they are likely to be smaller than the above-mentioned work suggests since the cortical dominance observed is likely due to the bias of MEG to superficial brain tissue.

Regardless of the actual relative neural contributions, one immediate consequence that results from this emerging view of the FFR's origins is that any terminology that seems to imply exclusive brainstem origins ought to be avoided in future publications for the sake of clarity and accuracy. Many of the more recent articles that display sensitivity to this issue seem to simply refer to the 'FFR', or some more specific instance such as the 'human FFR' or the 'scalp recorded FFR'. The issue here is that the FFR is often used to ambiguously refer to either or both the envelope FFR and the spectral FFR! It is for this reason that Aiken and Picton (2008) distinguish between 'spectral FFR' (sFFR) and 'envelope FFR' (eFFR). These same two terms are adopted here for sake of clarity, and 'FFR' is used throughout to designate the broader concept that encapsulates both.

Genesis and nature of the FFR

Figure 2.2 depicts an amplitude-modulated tone (a carrier frequency of 1000 Hz modulated at 80 Hz) at two different levels – here labeled as 'soft' and 'loud' – as they move through the cochlea and get converted into neural signals in the auditory nerve. It is worth noting that the spectral properties of these tones vary only in the amplitude of their frequency components, and that they possess energy solely at their carrier frequency and the two side-

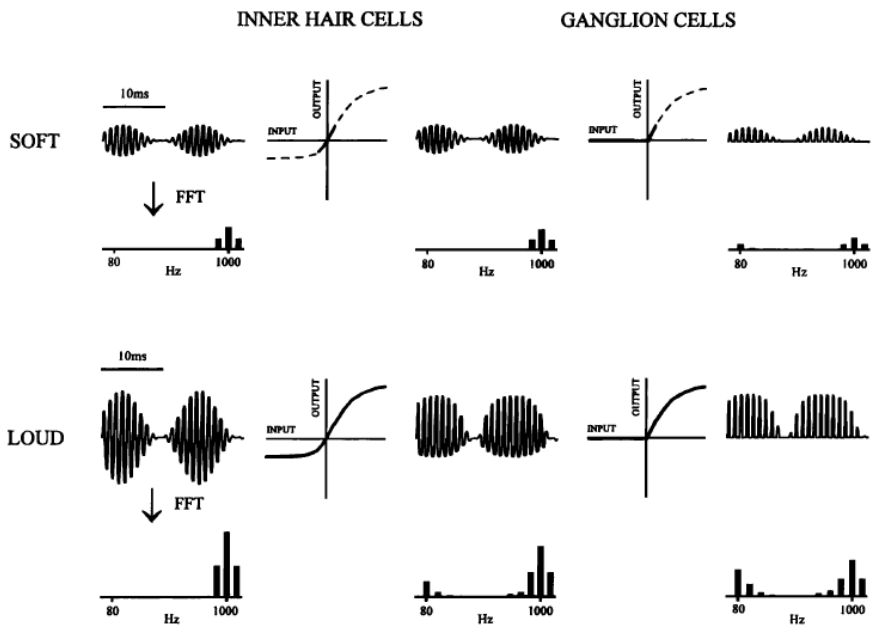

Figure 2.2: The generation of the spectral FFR and the envelope FFR that results when an am stimulus is transduced from its acoustic waveform (presented here at 'soft' and 'loud' levels) to electrical impulses in the auditory nerve. The spectra of the middle and right columns show energy at the modulating frequency of 80 Hz and its harmonics (eFFR components) and at the carrier frequency of 1000 Hz and its sidebands (sFFR components). Reprinted with permission from "Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz", by O. Lins et al. 1995, Journal of the Acoustical Society of America, 97, pg. 3059. Copyright 1995 Acoustical Society of America.

bands. The input-output curve of the inner hair cell (IHC) is roughly linear until the level is increased to a point where a noticeable compression occurs. This effect can be seen in the middle column, as the processing of the 'soft' tone appears to be roughly the same as its input, whereas the output of the 'loud' tone shows significant and asymmetric compression (due to the asymmetrical behaviour of the hair-cell – see Appendix A). The so-called 'half-wave' rectification of the signal at the level of the afferent auditory nerve (AN) fibres can be seen in the final column. The spectrum of the response to each tone clearly shows that there is now power present at 80 Hz (the modulation frequency) and its harmonics. The tracking of these frequency components throughout the auditory system underlies what we refer to as the envelope frequency following response. Similarly, the response to the carrier frequency and its side bands contributes to what we refer to here as the spectral frequency following response.

Regardless of terminological preferences, it should be clear that the FFR is a good measure of the phase-locking activity to the temporal fine structure of an acoustic signal, as well as to its envelope. Due to the robust and stereotyped nature of the response across a large variety of stimuli (e.g. clicks, tones, tonal sweeps, formant transitions, speech sounds, musical intervals, iterated ripple noise, etc.), it is unsurprising that it has not only been used to estimate thresholds, but that deviations from the normal population response can often

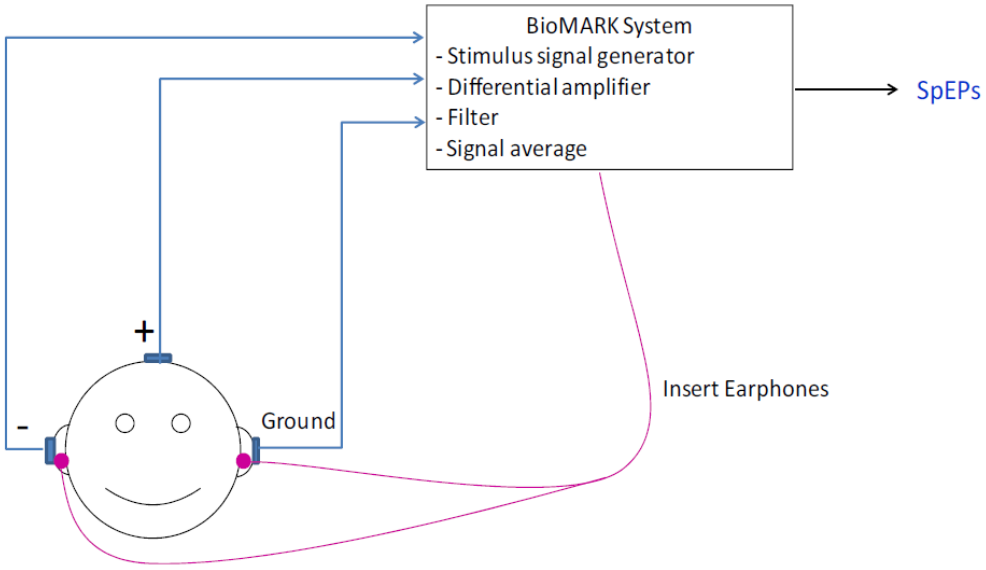

Figure 2.3: Schematic depicting a typical FFR recording setup, with the active electrode (+) placed at Cz, and reference (-) and ground electrodes on opposing earlobes. In many cases, the stimulus is only presented to one ear (typically the right ear). Reprinted with permission from Classification of Speech Evoked Responses To English Vowels (p. 29), by A. Sadeghian, 2012, Ottawa, Canada: University of Ottawa. Copyright 2012 Amir Sadeghian.

be used as an indicator of language impairment, literacy, dyslexia, autism spectrum disorder, musical experience, bilingualism, presbycusis, hidden hearing loss, and other manner of behavioural advantage or disadvantage where differential auditory abilities are expected or suspected to play a role (Kraus, Anderson, & White-Schwoch, 2017). Given that the FFR can be recorded with or without active engagement from the subject, that it is relatively easy to set up, and that it is a good indicator of auditory processing abilities in general, with specific indicators for speech-processing abilities, it seems clear that is worthy of consideration in an automated hearing aid fitting paradigm.

2.2.1 So how do we capture the FFR?

The FFR is most typically recorded using a simple electroencephalographic (EEG) montage when the subjects are human. Invasive electrode recording techniques are feasible, and are often used when recording animal models, but the bulk of the literature related to human subjects results from a simple 3-electrode setup: active, reference, and ground (see Fig. 2.3). MEG is also used to capture the FFR, although it appears much less frequently in the literature and is not typical of any clinical applications to my awareness.

With respect to the speech-evoked FFR, the overwhelming majority of studies record from an active electrode placed at Cz, with reference and ground recorded from opposing earlobes or mastoids. It should be noted that this is not always the case, however and that the choice of sites for electrode placement can and often does significantly alter the results.

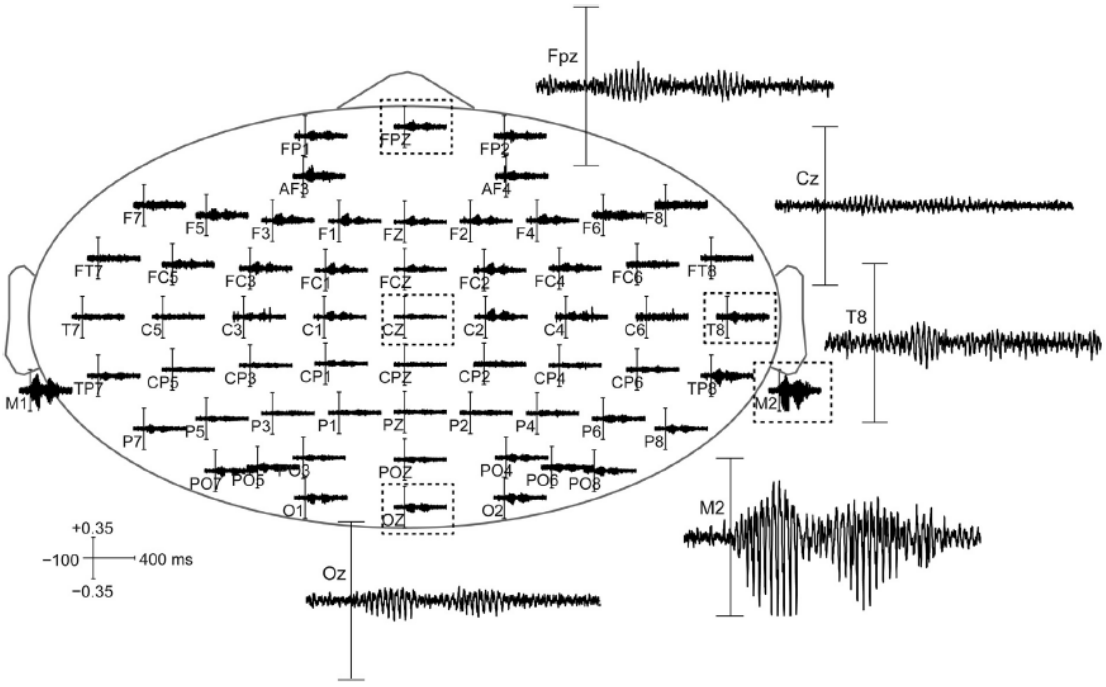

Figure 2.4: The impact of recording location on the speech-evoked FFR (/ama/, 300 ms). Responses were recorded at 64 locations around the scalp using an average of all of the channels as a reference. Reprinted with permission from "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from transient ABR", by G. Bidelman, 2015, Hearing Research, 323, p. 73. Copyright 2015 by Elsevier B.V.

Figure 2.4 is due to an excellent study by Bidelman (2015) on the effects of recording site and the various neural generators that contribute to the FFR, and it shows the resulting spectral FFR elicited from the same response recorded at each of the international 10-20 system locations devised for standardizing EEG recording and clinical practices.

The reason for these differential effects of recording location is simple: the FFR is comprised of multiple neural generators, each with its own dipole orientation. Therefore, what we record as the FFR is ultimately the sum of multiple overlapping waveforms that add and cancel one another in complex ways depending on the frequency, phase and amplitude of components emitted from each dipole, as well as the conductive properties of the brain matter that lies between said dipoles (Tichko & Skoe, 2017).

As mentioned in the section above, there is still no general consensus as to exactly which neural sources are contributors to the FFR, and in what varied proportions depending on recording location and the properties of the stimulus employed to evoke it. However, there is general agreement that the FFR is indeed a neural response, and that likely contributors include some combination of AN, the lateral lemniscus (LL), the cochlear nucleus (CN), the inferior colliculus (IC), the medial geniculate body (MGB), and the auditory cortex (AC) (see Appendix A for descriptions). Beyond electrode placement, the electrode material and the various gels and pastes that are often use in order to secure a good connection with the

scalp, as well as the thickness of the scalp, any oils or products, hair, etc. all contribute to the overall impedance at each electrode, and ultimately introduce additional sources of variability into a recording.

2.2.2 FFR study-related parameters

A great deal has been written on the methodological considerations related to the recording of the FFR, and the reader is again referred to the tutorial on the subject by Skoe and Kraus (2010), as well as to the two very thorough texts dedicated to the broader study of auditory evoked potentials by Picton (2010) and Burkard et al. (2007) for details on all considerations. A tabulated summary of the general recommendations related to stimulus type, duration, intensity, and polarity, as well as aspects of stimulus presentation and recording parameters related to electrode montage and materials, sampling rate, filtering, signal averaging, and artifact minimization are given in Appendix C.

Briefly, the choice of stimulus is dependent upon the aims of the study or assessment, and so for instance, musical intervals might be presented in order to analyze the correlates of musical consonance, or speech sounds might be used in order to diagnose auditory processing disorder. The presentation level of the stimulus is almost always well above detection threshold but held within a comfortable range (i.e. 55-85 dB SPL), as responses near to threshold are often weak and would require many repeated sweeps to yield a sufficiently strong signal (see Fig. 2.5). As the recording time of an experiment or an assessment is always a necessary consideration, the duration of the stimulus is often as brief as is possible while enabling the accurate capture of both the transient and steady state responses to components of interest. For speech stimuli, this typically results in tokens of around 40-300 ms. The fundamental frequency (F0) of any periodic (or quasi-periodic) stimulus is often maintained somewhere between 80-500 Hz, as this is typical of the human speech range and is well within the upper limits of phase-locking in the auditory nerve and throughout the brainstem, which has an upper bound of roughly 1500-2000 Hz. The stimulus can be presented either through earphones or through speakers (free-field), with insert earphones often being preferred over supra-aural earphones when appropriate in order to minimize stimulus artifact (due to the transduction occurring at a distance from the actual insert, separated by tubing) and to ensure consistent presentation level.

Due to the high temporal sensitivity of the FFR, specialized hardware and software is required in order to ensure that there is no jittering in the stimulus presentation. This is a requirement resulting from the fact that the response is on the order of microvolts, and the FFR is therefore only revealed as the result of signal averaging, often formed of two or

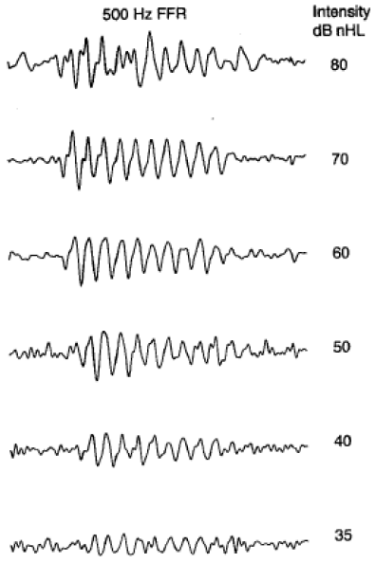

Figure 2.5: Tone evoked eFFR at different levels. sFFR waveforms in response to a 500 Hz pure-tone stimulus presented at different stimulus levels across the speech range. Notice the generally increasing amplitude and slightly decreasing latency with increasing level. In this case, the sFFR amplitude asymptotes around 70 dB SPL, and then decreases with further increases in presentation level. Reprinted with permission from Auditory Evoked Potentials: Basic Principles and Clinical Applications (p. 315), by R. Burkard et al., 2007, Baltimore, MD: Lippincott, Williams, & Wilkins. Copyright 2007 by Wolters Kluwer Health.

more sub-averages of 1000-3000 sweeps. Thus, if there is jitter on the order of tenths of milliseconds or greater either on the presentation or recording end of the system, then the signal is likely to be distorted or entirely ablated during averaging.

Once the stimulus has been selected, and the presentation system and the associated parameters are properly setup, only a few other recording considerations remain. The amplification and filter settings must also be set. Each response passes through an amplifier, with a typical gain in the range of 10,000-200,000, often a notch filter at 60 Hz (or 50 Hz depending on AC mains in the region), and a bandpass filter with a low cutoff between 30-100 Hz and a high cutoff between 1000-2000 Hz, depending on the frequency regions of interest. It is then important that the sampling rate of the recording system be set such that it is at least at twice the highest frequency of the signal in order to avoid aliasing. In modern practice, most systems oversample at rates around 20 kHz, which far exceeds the aforementioned upper frequency of the signal after passing through the bandpass filter. Doing so ensures a high degree of temporal precision and removes any possibility of misrepresenting the frequency contents of the response. Finally, the inter-stimulus interval (ISI), typically in milliseconds, or alternatively the stimulus rate, typically in presentations per second, must be selected. Ideal selection criteria here involve ensuring that the response has returned to baseline before any subsequent presentation, and that there are minimal perceptual distortions associated with the presentation rate, both of which are to be accounted for while simultaneously trying to minimize the duration of the recording procedure (Skoe & Kraus,

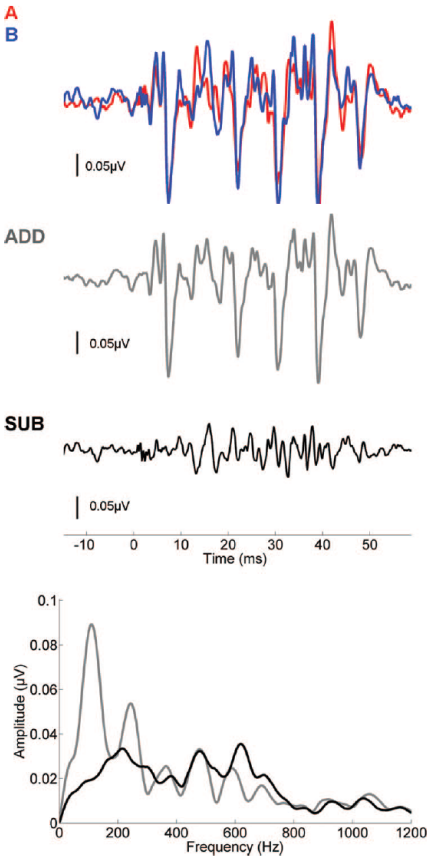

Figure 2.6: Stimulus polarities, and eFFR and sFFR. The EEG recording of the two polarities of a 40 ms /da/ stimulus are shown in red (A) and blue (B) in the top panel. Both are visibly similar, especially with respect to the most prominent peaks that occur at intervals of the period of the fundamental frequency. Adding both polarities results in the envelope FFR (ADD); subtracting them yields the spectral FFR (SUB). The spectrum in the bottom panel clearly shows the enhanced tracking of the lower frequency components present in the temporal envelope (gray), which is characteristic of the envelope FFR, whereas the spectral FFR (black) exhibits enhanced high-frequency components related to the vowel formants. Reprinted with permission from "Auditory Brain Stem Response to Complex Sounds: A Tutorial", by E. Skoe and N. Kraus, 2010, Ear and Hearing, 31, p. 311. Copyright 2010 by Lippincott Williams & Wilkins.

2010).

Older systems tend to generate the FFR sub-averages while discarding individual responses, and so they employ automatic artifact rejection, with a ceiling of anywhere from 20-50 microvolts. Newer systems typically record and store data from each and every electrode during every sweep, allowing for offline re-referencing and rejection, as well as the determination of important metrics based on individual sweeps, such as the phase-locking value. Most FFR recordings make use of alternating polarities (condensation and rarefaction, achieved simply by multiplying the original stimulus by -1), and often aim to record half of each sub-average in each polarity, thereby enabling the calculation of both the spectral and envelope FFRs (see Fig. 2.6 for an explanation of how this is achieved, and Table 2.2 for a breakdown of the different constituent components of each of the possible sub-averaging combinations of the two polarities).

Finally, as concerns the preparation of the subject, once the electrode montage has been determined, the subject's skin is typically prepped by using a mildly abrasive skin prep pad, followed by the wiping away of dead skin, loose hair, and scalp oils via an alcohol swab, and then by the application of a skin prep gel to the scalp recording site, and lastly by the application of a conductive electrode paste to the selected electrodes prior to their placement

Table 2.2: Average response nomenclature. A summary of the possible combinations of the responses to an original stimulus and its inversion (i.e. its opposite polarity), along with the components contained within the signal as formed. Reprinted with permission from "Envelope and spectral frequency-following responses to vowel sounds", by S. Aiken and T. Picton, 2008, Hearing Research, 245, p. 36. Copyright 2008 by Elsevier B.V.

Response	Derivation	Components
++	Average together all responses to the original stimulus	Envelope FFR Spectral FFR Cochlear microphonic Stimulus artifact
+−	Average together an equal number of responses to the original stimulus and responses to the inverted stimulus	Envelope FFR
−−	Subtract responses to the inverted stimulus from an equal number of responses to the original stimulus and divide by the total number of responses	Spectral FFR Cochlear microphonic Stimulus artifact

in the case of reusable electrodes, or simply by the placement of the disposable electrode once the skin prep gel has dried sufficiently. The impedance of all of the electrodes is then measured, with a typical acceptable range of less than 6 kΩ.

Depending on the nature of the study, the subject either receives an explanation of the active tasks to be performed, or is simply instructed to recline comfortably and to disregard the stimulus to the extent possible in the case of passive design. In either case, it is important that the subject is able to minimize any unnecessary muscle activation, as myogenic artifacts are a common source of signal contamination, and often result in rejected sweeps, which can substantially increase the recording time of a session. It is very common for a film to be played, often on silent for adults, but commonly at quiet level (~45 dB SPL) for child participants, in order to reduce fidgeting and to control for their cognitive state, as it is not uncommon for participants to exhibit symptoms of drowsiness during extended recording sessions due to what is often a monotonous procedure occurring in a dark, quiet and isolated environment.

2.3 The effects of level and amplification on the FFR: key papers

Only a few studies that directly assess the effects of level on the speech-evoked FFR have been published to date. In fact, at the commencement of this research, only 3 studies that clearly focused on the matter had been published, and none of these performed anything resembling a comprehensive analysis of both the envelope and spectral FFRs in both of the time and frequency domains. Very recently, some additional results have been published, and almost all of the research was undertaken with similar aims to those proposed herein,

or at least in part, as they sought to record and analyze the FFR in HI populations in aided or unaided conditions with therapeutic applications in mind.

A summary of the studies most relevant to the research described herein is provided in Table 2.3 in order to simultaneously highlight what has been done to date, as well as to reveal the gaps in the literature in order to demonstrate how my results might serve to fill some of them in. Note that only FFR-based studies are included here, and additional classes of AEPs such as CAEPs have been excluded. A subsequent review of the application of the broader class of auditory evoked potentials in hearing aid fitting and verification is provided in Section 3.3, again with a similar analysis of work that remains to be done.

Prior to commencing, it is worth noting that a conference paper by Dajani, Heffernan, and Giguère (2013) provides an overview of our thinking on the subject of the FFR's utility in HA fitting in addition to some preliminary data. This paper, along with a short letter (Kraus & Anderson, 2012) and a published review (Anderson & Kraus, 2013) were all submitted around the same time, and serve as an early conceptualization of the space that we seek to delineate.

Table 2.3: Key papers presented in chronological order, with descriptions of the stimuli used, the presentation level(s), subject population(s), and the type(s) of analysis performed.

Paper	Stimuli	Level(s)	Subjects	Analysis			
				Time		Freq	
				eFFR	sFFR	eFFR	sFFR
Krishnan (1999)	/u/, /ɔ/, /a/ two-tone approx. of steady state vowels; 60 ms each	55, 65, 75, 85 dB nHL	10 NH adults	N	Y	N	Y
Plyler and Ananthanarayan (2001)	100 ms /ba/, /da/, /ga/; F0: unspecified F1: 150 → 750 Hz; F2: /ba/: 900 → 1200 Hz; /da/: 1600 → 1200 Hz; /ga/: 2300 → 1200 Hz F3: 2300 Hz	72, 82, 92 dB SPL	16 NH adults; 16 HI adults	N	N	N	Y
Krishnan (2002)	3 synthetic steady state vowels: /u/, /ɔ/, /a/; 60 ms each	55, 65, 75, 85 dB nHL	10 NH adults	N	Y	N	Y
Akhoun et al. (2008)	/ba/; 50 ms; F0=120 Hz; F1=700 Hz; F2=1200 Hz	0-60 dB SL in 10 dB steps	23 NH	Y	N	Y	N
Fu, Liang, Su, and Wang (2009)	/da/; 40ms; F0~100 Hz	60 & 80 dB SPL	31 young Chinese adults (17 F)	N	Y	N	N
Kraus and Anderson (2012)	170 ms /da/; F0=100 Hz	Two amp. settings; unknown level	1 person case study	Y	N	Y	N
Anderson and Kraus (2013)	As above	As above	As above	Y	N	Y	N

Anderson, Parbery-Clark, White-Schwoch, Drehobl, and Kraus (2013)	40 ms /da/; F0~103 → 125 Hz; F1: 220 → 720Hz; F2: 1700 → 1240 Hz; F3:2580 → 2500 Hz	80.3 dB SPL for both groups, and amplified for each HI subject per individual thresholds	30 adults: 15 NH, 15 HI	N	N	Y	Y
Dajani et al. (2013)	300 ms /a/ stimulus; F0=100 Hz; F1=700 Hz; F2=1200 Hz	55, 65, 75, and 85 dB SPL	1 NH young adult	N	N	Y	Y
Bellier et al. (2015)	/ba/ - 180 ms natural French (F) voice with /a/ having F0 ~170 Hz, F1~740 Hz, F2~1360 Hz, F3~785 Hz, F4~4310 Hz	80 dB SPL across three conditions: insert earphones, and HAs at two settings	4 NH adults (native French speakers)	Y	Y	Y	Y
Easwar, Purcell, Aiken, Parsa, and Scollie (2015a)	/susaji/ - naturally male with average F0: 96-100 Hz; 2.05 s; /a/: F1:684 Hz; F2:1467 Hz; 447 ms; /u/: F1:266 Hz; F2:1016 Hz; 386 ms	Exp 1: 50 and 65 dB SPL; Exp 2: 65 dB SPL, low-pass filtered at 1, 2, and 4 kHz	20 NH young adults	N	N	Y	N
Easwar, Purcell, Aiken, Parsa, and Scollie (2015b)	As above	Exp 1: 50 and 65 dB SPL in unaided conditions and through individually verified HAs; Exp 2: 65 dB SPL, low-pass filtered at 1, 2, and 4 kHz	21 adult HA users with mild to moderately severe SNHL	N	N	Y	N
Ananthakrishnan, Krishnan, and Bartlett (2016)	/u/: 265 ms synthetic steady state with F0=120 Hz, F1 = 360 Hz, F2=970 Hz, F3=2667 Hz, and F4=3007 Hz	Exp. 1: 60-85 dB SPL in NH listeners and 70-95 dB SPL in HI listeners (in 5 dB steps); Exp. 2: 80dB SPL	Exp. 1: 10 normal hearing; 9 with mild-moderate SNHL; Exp. 2: 25 NH; 19 with mild-moderate SNHL	N	N	Y	Y
Hassaan, Ibraheem, and Galhom (2016)	/ba/, /ga/, and /da/ at 114.875 ms, 213.250 ms, and 206.275 ms, respectively; vowel portion in range: F0~99-112 Hz; F1~730-820 Hz; F2~1335-1420 Hz; F3~2025-2500 Hz	60 dB nHL	40 adults with bilateral SNHL ranging from moderate to severe; 30 ears without dead regions; 46 with dead regions at low, mid, or high frequencies	Y	N	N	N

The two most comprehensive studies performed exclusively on normal hearing subjects were done by Krishnan (1999, 2002). They are both extremely similar in design, with both studying 10 normal hearing subjects across the same four levels (55, 65, 75, and 85 dB nHL); however, the first uses a very simplified set of stimuli, namely two-tone steady state approximations to three vowels (/u/, /ɔ/, and /a/), whereas the second actually uses a fuller synthetic steady state version of the same. Both studies also borrow from a nearly identical analytical framework which is limited solely to the analysis of sFFR in the time and frequency

domains. The results generally indicate strong phase-locking and robust formant capture via a non-decreasing (i.e. increasing and then saturating) effect on the spectral amplitude of both the first and second formants in addition to the presence of some distortion products, as can be seen in Figs. 2.7 and 2.8 (see Appendix D for the results of the remaining vowels).

Unfortunately, little is said with respect to the time domain waveforms other than to remark their consistent tendency to increase in amplitude with increasing level, and to note that certain vowel stimuli result in relatively more complex waveforms than do others. No actual measures of this tendency are given, nor are any measures of latency given, nor are there any attempts to characterize any onset or offset responses or any other associated temporal markers. It is worth noting that the results from the 1999 paper indicate the general dominance of the first formant (F1) in the neural response, as well as the possible contribution of two-tone suppression, whereby F1 is suppressed by the second formant (F2) at increasing levels, as the fibres with characteristic frequencies (CFs) between F1 and F2 that would otherwise contribute to the encoding of F1 are increasingly phase-locked to F2.

Another study from this time period by Plyler and Ananthanarayan (2001) examined the effects of the second formant transition in both normal and impaired hearing groups, and at three different levels of 72, 82, and 92 dB SPL. They assessed the changes in spectral content of the sFFR from the transient to the steady-state portion of the stimuli in order to quantify the encoding of the second formant transition, and observed no systematic shift in the frequency peak over time for any stimuli at any level in the hearing impaired group. In contrast, the normal hearing group demonstrated such a shift corresponding to the temporal changes in the second formant frequency for each stimulus (although presumably not for each level, as no such claim was made, nor was any such effect or lack thereof explicitly addressed). The authors suggest that these differences might be accounted for by the diminished phase-locking capabilities of the hearing impaired group.

The most comprehensive study of the time domain effects of speech level is by Akhoun et al. (2008). Using a 40 ms /ba/ syllable beginning with a stop consonant at 10 dB steps from 0 dB SL to 60 dB SL, characteristic peaks and their apparent lags were used in order to characterize both the onset response and the steady-state portion of the eFFR, which are clearly seen in Fig. 2.9. Notice the very consistent decrease in latency with increasing level across all conditions, and the difference in slope between the decreases in the peaks representing the onset response in comparison to those representing the steady state response (here labelled FFR). Another study by Fu et al. (2009) that assessed the temporal effects of stimulus level using a 40 ms /da/ at 60 and 80 dB SPL also observed a decrease in latency and an increase in amplitude across many waveform features with an increase in level. They also observed greater correlations between these same waveform features at the lower stimulus

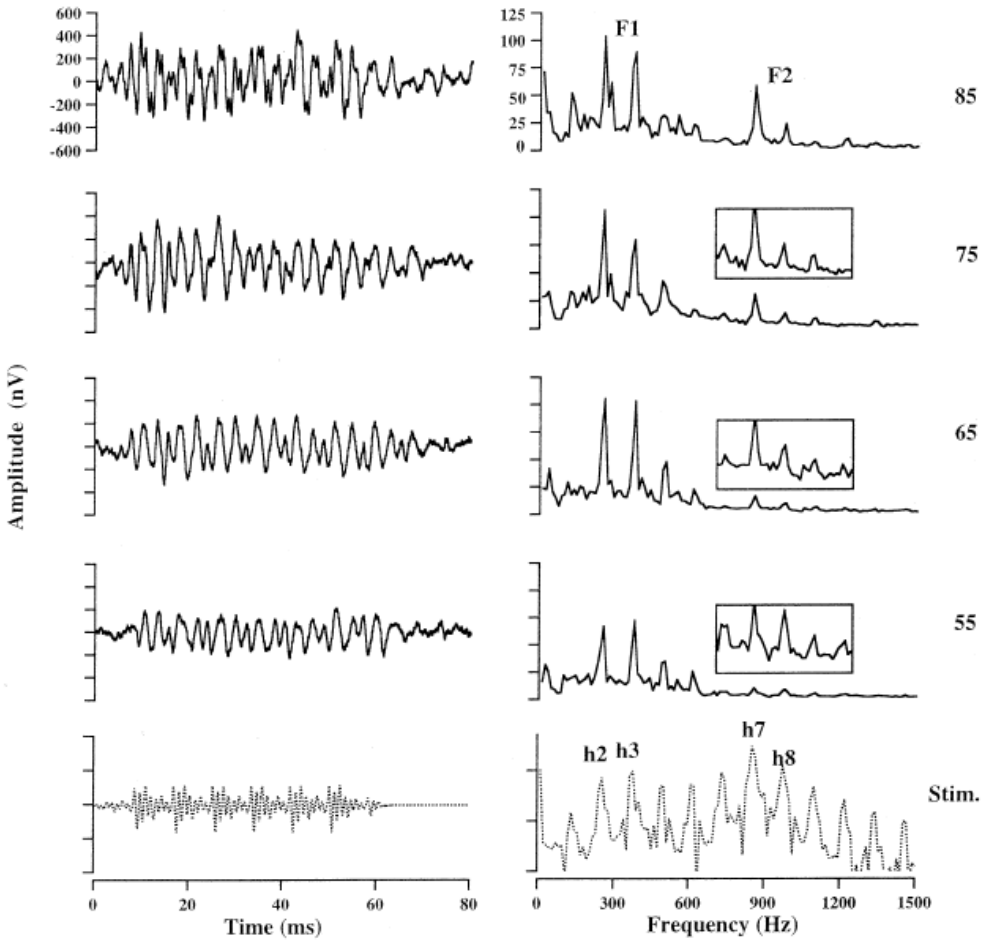

Figure 2.7: Grand-average sFFR for /u/ vowel. The grand average sFFR waveforms (left panel) and spectra (right panel) plotted as a function of sound intensity (dB nHL) for the /u/ vowel. The stimulus waveform and its spectrum, along with the harmonics of F1 (h1, h3) and F2 (h7, h8) are shown along the bottom row. The insets show the F2 harmonic peaks present at all levels. Reprinted with permission from "Human Frequency-Following Responses: Representation of Steady-State Synthetic Vowels", by A. Krishnan, 2002, Hearing Research, 166, p. 194. Copyright 2002 Elsevier Science B.V.

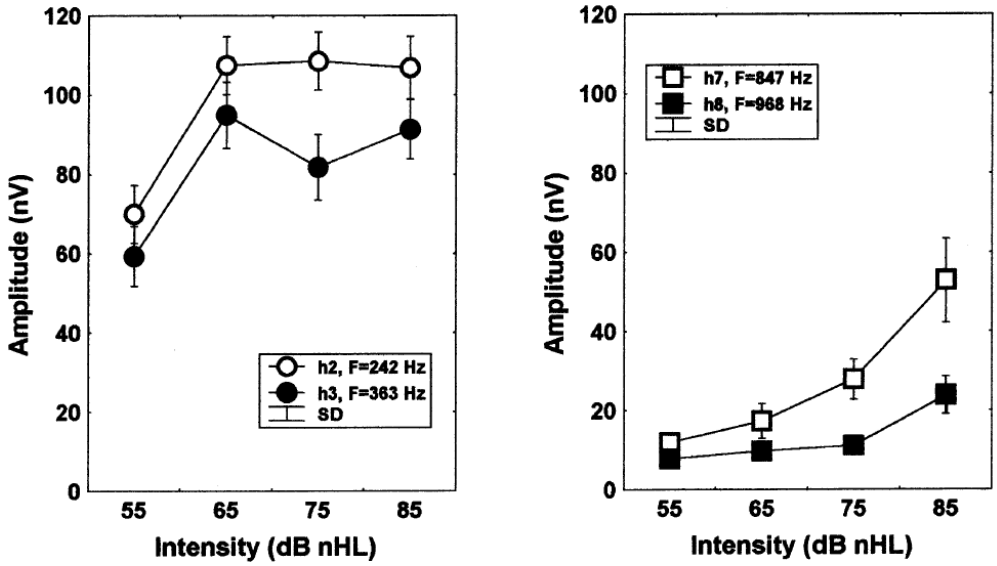

Figure 2.8: Mean amplitudes for the F1 (left) and F2 (right) response components in the sFFR, plotted as a function of sound intensity (dB nHL) for the same /u/ vowel stimulus in Fig. 2.7. The error bars indicated ±1 standard deviation (SD). Notice the general increase in amplitude, and the saturation that clearly occurs in amplitudes of the F1 harmonics. Reprinted with permission from "Human Frequency-Following Responses: Representation of Steady-State Synthetic Vowels", by A. Krishnan, 2002, Hearing Research, 166, p. 194. Copyright 2002 Elsevier Science B.V.

level. Neither of these studies endeavoured to assess their data in the frequency domain.

Save for two, all of the remaining papers tabulated above deal with either the impacts of hearing impairment or hearing aid usage on the FFR, or both. One among the two belongs to a pair of papers by Easwar et al. (2015a, 2015b) that assessed the envelope FFR to a complex natural speech token that is presented to both normal and impaired populations in unaided and aided conditions at 50 and 60 dB SPL to study the effects of level, and at 60 dB SPL (strictly in an aided condition for the HI subjects) for low-pass filtered versions of the same stimulus with cutoffs of 1, 2, and 4 kHz to study the effects of bandwidth as well. Due to the alterations made to the stimulus, whereby the modulation of the carrier representing the first formant of each vowel was lowered by 8 Hz from the naturally occurring F0 of ~98 Hz, a total of 8 responses were possible in each condition. In normal hearing subjects, the results showed both a significant increase in the amplitude and the number of detected (i.e. statistically significant) responses for all normal subjects with increasing level. A similar result was obtained for HI subjects: the increase in level resulted in a significant increase in the number of responses detected per condition, and at positive sensation levels (i.e. levels at which sounds were perceived by subjects), an increase in level resulted in a significant increase in amplitude in both aided and unaided conditions. The amplitude of the responses and the number of detections generally increased with increasing bandwidth in both groups, as did scores on behavioural tasks that were administered. Importantly, results indicate significant positive correlations between the behavioural test scores and the envelope FFR parameters,

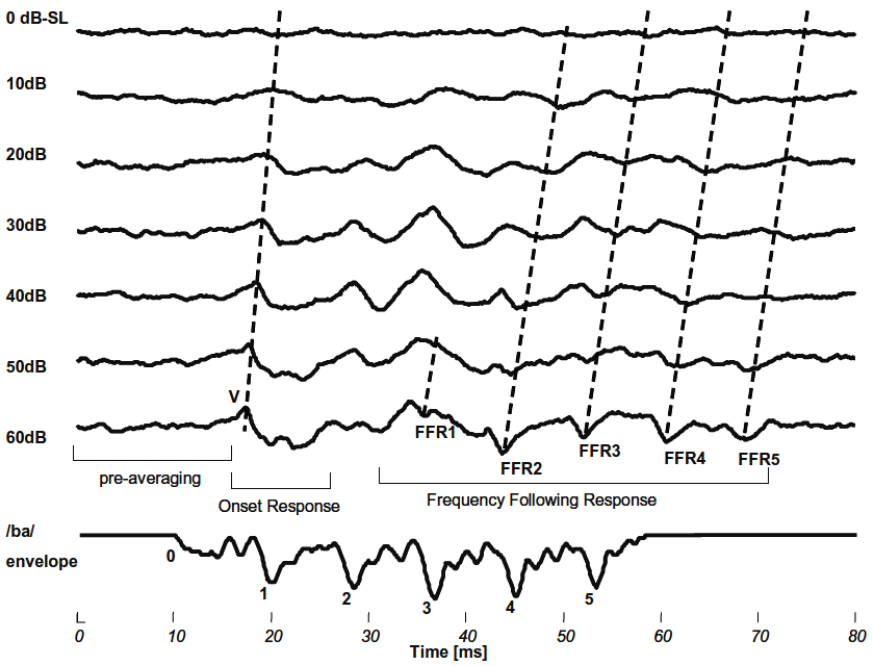

Figure 2.9: Envelope FFR latency/level function. Various features of the onset and envelope frequency following response (here simply labelled as FFR) are shown in relation to the increasing sound level of the /ba/ stimulus. The latency of all components denoted by a dashed line can clearly be seen to be decreasing with increasing level. However, the rate of change is different when comparing the onset response (V) to the frequency following response (FFR1-FFR5). Reprinted with permission from I. Akhoun et al., 2008, Clinical Neurophysiology, 119, p. 929. Copyright 2007 by International Federation of Clinical Neurophysiology.

which suggested that the paradigm may be of utility in an objective aided outcome measure.

The study by Bellier et al. (2015) focused on the development of methods for recording the speech-evoked FFR by stimulating directly through hearing aids via a wireless connection. At the time of its publication, the only prior study involving hearing aids was the case-study of a single subject by Anderson and Kraus (2013), which made use of open-field stimulation. Here, four normal hearing subjects were presented a 180 ms /ba/ stimulus in an unaided condition, and in two aided conditions with identical parameterization aside from the gain profile. The results indicated not only that artifact-free recordings are possible in this configuration, but that amplification tended to result in greater F0 amplitude in both the envelope and spectral FFRs as well. Interestingly, although not unexpectedly owing to the phenomenon of synchrony capture, the amplitudes of the upper harmonics tended to be reduced relative to the unaided condition. This resonates with the conjecture of Anderson and Kraus (2013) that optimal fitting might be one that somehow balances out the envelope and spectral FFRs, (i.e. the encoding of the envelope (ENV) and the temporal fine structure (TFS) – see Section 3.3 for a more detailed treatment of this subject).

An elegantly designed study by Hassaan et al. (2016) recorded FFRs from 30 subjects without cochlear dead regions, and from 46 subjects with dead regions at either low, mid,

or high frequencies. All subjects were recorded in a free-field and in an aided condition, and all had worn hearing aids for 5+ years. Three different stimuli were constructed so that their spectral maxima similarly occurred in these low, mid, and high frequency regions. By stimulating each subject with all three stimuli, they were able to show that the stimuli with spectral maxima in the dead regions elicited responses with longer latencies and smaller amplitudes in comparison to the control group or the responses to the other two stimuli. This is obviously very important to keep in mind when considering the construction of any BCI algorithm to optimize fitting. The authors further concluded that the impact of dead regions on chronic HA users can be minimized as a result of brainstem neuroplasticity and due to the redundancy of frequency component encoding that results from the upward and downward spread of excitation.

Finally, a study by Ananthakrishnan et al. (2016) looked at the effect of envelope and spectral FFR encoding in 10 normal hearing listeners, and 9 with mild-moderate sensorineural hearing loss (SNHL) when presented with a /u/ stimulus in 5 dB steps across levels from 60 to 85 dB SPL, and 70 to 95 dB SPL in the normal hearing (NH) and SNHL groups respectively. Their results show that NH subjects had stronger F0 (revealed via the eFFR) and F1 (revealed via the sFFR) responses at equal levels, and that comparison across SPL and sensation level revealed differential effects of level between both groups. A follow-up experiment that sought to determine the effect of age on the FFR was also performed, and it was observed that the degradation of both responses persisted in listeners with SNHL even when the effects of age were controlled for.

Importantly, this result adds some clarity to our understanding of the relative coding strength of the envelope and temporal fine structure in HI populations – a complex topic whose waters remain fairly muddied at present. Anderson and Kraus (2013), following along their proposed line of research that briefly outlined the potential utility of the FFR in managing HI (Anderson & Kraus, 2013; Kraus & Anderson, 2012), found greater eFFR responses in their HI population as compared to the NH group, and that the groups were equivalent with respect to the sFFR response. They in turn suggest that the relative coding strength between the two may be disrupted in SNHL, and that this imbalance with respect to the normal population may be the basis for the speech comprehension problems that are typically encountered in noisy environments. Ananthakrishnan et al. (2016) point out that this result is at odds with a previous result due to Plyler and Ananthanarayan (2001) that suggested a degradation in the temporal fine structure representation in the HI population with respect to second formant tracking as measured using the FFR. They suggest that this discrepancy between the studies might be due to the use of a single high presentation level by Anderson et al., as well as to age related effects common to both of their study

groups, caveats that Anderson et al. themselves offer up as a potential explanation for their somewhat surprising observations.

Although physiological studies of AN suggest that TFS is degraded in the HI population, while envelope responses remain robust, the relative strength of each is likely to vary depending on the speech token, being particularly dependent upon the formant structure for vowel tokens, as well as the subject's specific profile and the presentation level (note that AN fibres above the F2 region often synchronously discharge to the modulation rate, as outlined in Appendix A.4.3 and Fig. A.17). Regardless of this complex interplay, the majority of evidence supports a degraded TFS encoding due to decreased phase-locking in the periphery, and so it seems that the result of Anderson et al. should be taken with caution here. This is especially true when contemplating the design of any novel HA signal processing strategies, as Ananthakrishnan et al. (2016) point out.

2.4 A brief review of sex differences in the FFR

Sex is known to be a potential confound in AEP studies, with women tending to have shorter latencies and larger amplitudes than men – an effect which appears to persist from infancy through adulthood (Solanki, Mehta, et al., 2015). This seems to hold true for the click-ABR, the CAEP, and the speech-FFR, although some results are mixed, and the effects are often quite small, especially when sensible normalization techniques are employed (Dehan & Jerger, 1990; Jalaei, Zakaria, Mohd Azmi, Nik Othman, & Sidek, 2017; Nikiforidis, Koutsojannis, Varakis, & Goumas, 1993). Several potential factors have been proposed in order to explain these differences, including head and inner ear size differences which are believed to result in greater latencies due to longer travel times in anatomically larger males, as well as differences in efferent modulation and cortical structures and functions (Krizman, Skoe, & Kraus, 2012; Nikiforidis et al., 1993; Solanki et al., 2015). Interestingly, estrogen appears to account for some of the observed variance, as several studies show that females with diminished levels tend towards the longer latencies of males (Batta, Dhir, Kumar, & Singh, 2017).

After conducting a survey of globally sampled results, Solanki et al. (2015) suggest that setup-specific normative data for both sexes be established with a fairly large sample in order to maximize the utility of AEPs as a reliable tool whose results are more universally accepted. This is likely prudent, as a study on intersubject variability of click-ABRs found that variance values were reduced by approximately 40% when treating each sex separately, thereby enhancing the usefulness of latency as a parameter in characterizing auditory function (McClelland & McCrea, 1979). It is worth noting that typical differences in non-cortical latencies are on the order of 0.1-0.5 ms, which is often similar to the within-sex standard

deviations values (Hoormann, Falkenstein, Hohnsbein, & Blanke, 1992; Jalaei et al., 2017; Krizman et al., 2012; McClelland & McCrea, 1979).

With respect to the speech-FFR specifically, two studies have found that the components related to the onset response differed between sexes, with the latencies of females being shorter relative to males (Jalaei et al., 2017; Krizman et al., 2012). Krizman et al. (2012) showed that FFR latencies and spectral amplitudes to the fundamental frequency were not affected by sex; however, females possessed notably larger high-frequency elements in their representations. The authors suggest that this might account for the higher incidence of reading impairment in males, as poor readers are known to differ from good readers on encoding fast but not slow components of speech, although this should be taken with caution, as Tichko and Skoe (2017) recently showed that the fine-structure of the FFR as recorded at the scalp is frequency dependent due to constructive and destructive phase interactions between the various neural generators, and all of the above-mentioned sources of between-sex variability are likely to play a role here as well.

2.5 Gap analysis: towards a better understanding of level effects in the FFR

The background information and review of the literature reveals the following:

1. Almost all of the studies have a similar /a/ vowel in their stimulus or stimulus set, and typically in the same male-English frequency range

2. The duration of the stimuli, the bulk of which is comprised by the vowel portion(s), range in duration, with a bias towards being unnaturally short in duration item

3. None of the studies comprehensively analyses both the envelope and spectral FFRs in both of the time and frequency domains

4. All studies include data from both sexes, but none analyzes the results between groups despite well-known sex differences in the ABR

Generally, it can be seen that there has not yet been any attempt to broadly characterize the FFR response of normal hearing subjects to changes in level and duration across the conversational speech range, across vowel space (i.e. 2D formant space) and between the sexes. For the sake of analyzing the space of possible algorithms that might incorporate the FFR in any sort of automated fitting process, as is the stated purpose here, the creation of such a dataset seems to be a necessary stepping off point. And although some interesting

results on the impacts of hearing loss and amplification have recently been published, without a more comprehensive normative dataset from normal hearing adults for comparison, there are likely to be some relevant or exploitable features that go unused or unrecognized in the quest for a more optimal and efficient fitting process.

With respect to sex-level interactions, the bulk of the evidence seems to suggest that some small interaction effects might be observable in FFR studies of sex and level. This is obviously of interest from a pure research perspective, and is a very natural extension of previous work done to examine effects of sex in auditory processing. It is also just as obviously of interest from a practical research perspective, and regardless of the size of the effect in the general population and its diagnostic utility or the lack thereof, being able to significantly diminish unwanted variance in a dataset designed for practical ML and BCI purposes is certainly worth undertaking, as it would be very likely to decrease error rates and to increase classification accuracy as a result. This is especially likely to be the case when time-domain features are a part of the feature vector.

Chapter 3
Background: ML and BCI considerations related to HA fitting

3.1 A brief overview of BCIs, ML, and SFHAs

This past decade or so has seen an explosion in the field of brain-computer interfacing (Hamadicharef, 2010). Broadly, a brain-computer interface (BCI) is any system that utilizes neural signals to control a technological device, elicited either as a result of active or passive cognitive state changes. The current suite of BCI applications ranges from neuroprosthetics (e.g. controlling robotic arms), gaming applications, commercial meditation and wellness applications, spellers for persons with locked-in syndrome, and more. Although a lot of initial BCI technology met with limited success, the convergence of several fields currently undergoing exponential growth (e.g. neuroscience, machine learning, and sensor technology) has resulted in a great deal of very recent success across the BCI space.

Despite all of these advances, most BCIs rely on motor imagery or visual sensory input, and auditory-BCIs remain extremely rare. To date, no hearing-aid related BCIs exist on the commercial market, although large HA producers are certainly looking at possible approaches (Carlile, 2016), and Germany's interdisciplinary cluster of academic, medical and private research known as Hearing4All has a task group dedicated to developing BCIs for HAs (*Hearing4all - Task Group 7: Brain Computer Interface for hearing devices*, 2017).

Figure 3.1 depicts a typical BCI workflow. The two italicized schematic components are typical of a machine learning (ML) pipeline. ML is a field of computer science whereby computers are able to learn (or build models) without being explicitly programmed. Many recent advances in this field have yielded powerful algorithms that are able to optimize for the classification of different data inputs, or in the BCI case, neural states or signatures that are often highly correlated with some cognitive or perceptual state and can therefore be used to control a device in a meaningful and desired way. Since HAs amplify specific frequency channels, and often have multiple modes/algorithms (e.g. directional or omnidirectional microphone settings, wind reduction, etc.) that are either switched manually or automatically to optimize for comprehension depending on the user's environmental instance, we should be able to exploit any neural correlate of the user's perceptual abilities related to speech comprehension across these environments to tune the parameters of the hearing aid to improve

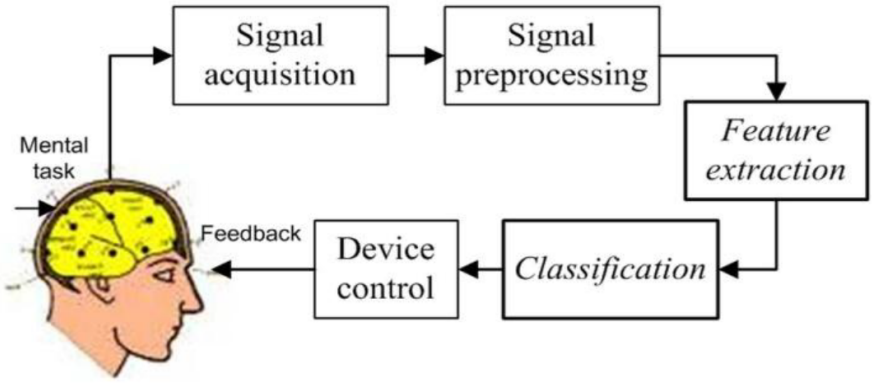

Figure 3.1: BCI schematic. The five main steps incorporated into typical BCI designs. Feature extraction and classification are particularly important in that the performance of the BCI depends on the accurate classification of the user state based on features of the EEG signal that are deemed to be relevant to the control process. Machine learning techniques are often used in the 'feature extraction' and 'classification' steps. Reprinted with permission from "P300 Detection Using a Multilayer Neural Network Classifier Based on Adaptive Feature Extraction", by A. Turnip et al., 2013, International Journal of Brain and Cognitive Sciences, 2, p. 63. Copyright 2013 by Scientific & Academic Publishing.

the fit of the hearing aid and to ensure its optimal performance throughout the lifecycle of the device.

The space of all potential BCI-HA constructions is quite large, and although some portion of this space is undoubtedly partitioned by assisted fitting and fine-tuning, whether by an audiological professional or not, the greatest foreseeable gains are reasonably envisioned to result from a self-contained self-fitting process. As mentioned, these SFHAs have the potential to address a large constituency of persons whose hearing health care needs are currently unmet due to prohibitive costs and limited access to the level of professional care demanded of conventional HAs (see Fig. 3.2 for a schematic breakdown of the SFHA space). The recent emergence of (non-BCI) SFHAs on the market has prompted renowned experts to predict a growth in this space, coupled with audiological assistance being offered through telehealth infrastructure that will be facilitated by the ubiquity of smartphones worldwide, regardless of the developmental status of a region.

3.2 A more in-depth look at machine learning

Machine learning has burgeoned into a huge field over the last decades, but it can be broadly described as the act of automatically learning programs or models from data (Domingos, 2012). Three broad categories of ML tasks exist: supervised learning, unsupervised learning, and reinforcement learning (Witten, Frank, Hall, & Pal, 2016, Ch. 1). Reinforcement learning involves learning a program that interacts with a dynamic environment (e.g. self-driving cars), and the programs improve or 'learn' via feedback and numerical 'punishments and rewards' (this evaluative metric is more commonly referred to as the objective function

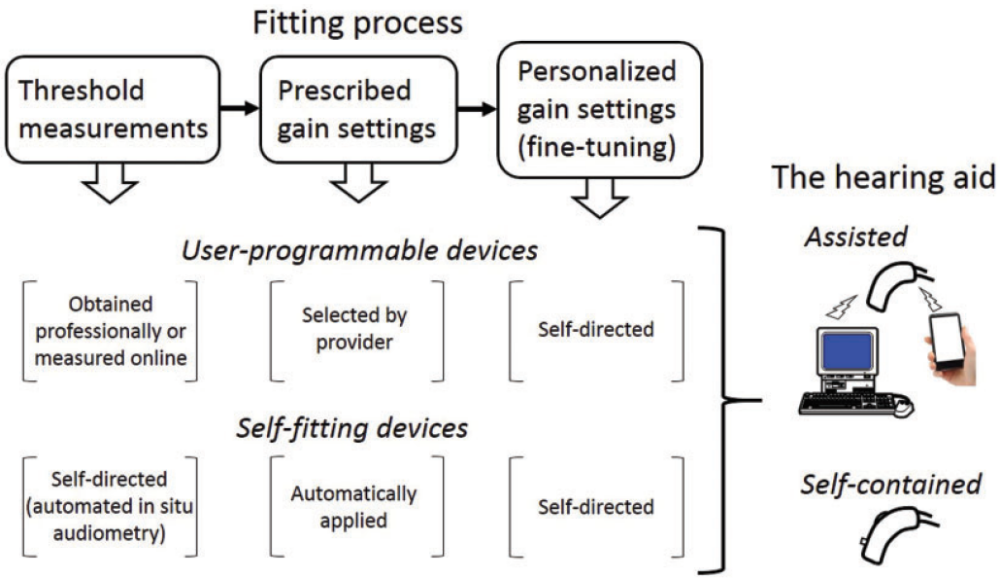

Figure 3.2: Self-fitting hearing aid schematic. An overview of the self-fitting process broken down by devices that can be described to be user-programmable or truly self-fitting. The process can be managed either via software, in an 'assisted' fashion, or it can be entirely self-contained, with all of the necessary software being integrated into the device itself. At present, all commercial devices are of the assisted variety. A variety of BCIs can be envisioned, whereby the prescribed gain settings and/or fine-tuning can be set as a result of the processing of relevant neural signatures associated with auditory processing of one variety or another. Reprinted with permission from "Self-Fitting Hearing Aids Status Quo and Future Prediction", by G. Keidser and E. Convery, 2016, Trends in Hearing, 20, p. 2. Copyright 2016 by G. Keidser and E. Convery.

or the cost function). Unsupervised learning involves learning from unlabeled data, and often involves finding hidden structure in datasets, or in learning features that might not be obvious to humans, and that can themselves be used in a supervised learning task. Supervised learning involves presenting an algorithm with a variety of labeled inputs, so that a model that generalizes well across the entire input space can be learned. Due to the current limitations in the size of datasets, and the fact that virtually all known FFR studies involve the capture of labeled data (i.e. the subject characteristics are known, the stimulus and recording conditions are known, etc.), the focus here will be on supervised learning, as all of the ML work presented here is of the supervised nature.

The most common end-goal of ML algorithms is that of classification, although it is certainly not uncommon for end-users to be interested in regression and clustering tasks, among others (here again the discussion is restricted to the discussion of classifiers, as they are the sole focus of the ML work presented here) (Domingos, 2012). A classifier is a model that receives a feature vector as an input, and outputs a single discrete value that corresponds to the predicted class of said input. In order for a useful model to be created, learning, evaluation and optimization procedures must be performed (Domingos, 2012). The former two are commonly performed by parsing datasets via a method known as train/test split, wherein roughly 50-80% of the available data is utilized to train the model, and the remaining 50-20% is held out in order to validate and test the model's performance along some sensible

metric (e.g. classification accuracy) (Witten et al., 2016, Ch. 5). The latter – optimization – involves algorithms for finding optima, with gradient descent being a common introductory method. This method is in some sense analogous to only taking small downward steps until you find yourself in a valley (i.e. a local or global minimum).

Training is accomplished by feeding examples (x_i, y_i), where $x_i = (x_{i,1}, x_{i,2}, , x_{i,d})$ is an observed input with d features (e.g. the amplitude of the first d harmonics in the spectral FFR), and y_i is the corresponding class value (e.g. 85 dBA /u/ vowel) (Domingos, 2012). The testing then involves inputting the remaining examples and comparing the output of the model (i.e. the predicted class) to the actual known class label. A simple metric of performance then can be the calculation of how frequently the difference between the predicted class and actual class (which are assigned numerical values in the model – e.g. 85 dBA = '4', say) is zero. Since it is obviously possible to have biased sampling of training and testing data, and the major concerns are to the appropriate generalizability of the learned model, it is often common to employ a technique known as cross-validation, that will split the data into k folds (k=10 is common), effectively repeating the train/test split procedure over ten different samplings of the dataset, and thereby allowing for a more honest assessment of the model's generalized performance (Witten et al., 2016, Ch. 5). For very small datasets, a training and testing method known as leave-one-out (LOO) is often used, wherein one single case is held out while a model is trained on all of the remaining data before predicting the class of the sole remaining example, and this procedure is repeated for all of the data points available (Witten et al., 2016, Ch. 5).

As Table 3.1 indicates, there are a variety of possible representations and associated classification models, as well as a variety of different ways to evaluate and optimize them (Domingos, 2012). Some algorithms work by splitting some aspect of the feature space by creating linear or nonlinear boundaries (e.g. support vector machines), whereas others such as the k-nearest neighbour algorithm simply looks at the k-nearest labelled points and proceed to label the input according to the most common class among them (Witten et al., 2016, Ch. 4). Aside from accuracy and error rate, it is common to see precision and recall used as metrics (Witten et al., 2016, Ch. 5). It is common for classification tasks involving supervised learning to make use of a confusion matrix, that enables the visualization of the performance of an algorithm by displaying a matrix of bin counts of the actual classes versus the predicted classes, and thereby the number of times that each class is accurately predicted, as well as inaccurately predicted to be one of the other classes (i.e. the model is confusing two classes) (Witten et al., 2016, Ch. 5).

Since there are a variety of freely available high-quality ML libraries, a large aspect of ML in practice amounts to appropriate feature and model selection, and model parameterization

Table 3.1: The three components of machine learning algorithms. As can be seen, there are a variety of different representations and associated models, methods of evaluating performance, and approaches for optimizing models. Reprinted with permission from "A Few Useful Things to Know About Machine Learning", by P. Domingos, 2012, Communications of the ACM, 55, p. 80. Copyright 2012 ACM.

Representation	Evaluation	Optimization
Instances	Accuracy/Error rate	Combinatorial optimization
K-nearest neighbor	Precision and recall	Greedy search
Support vector machines	Squared error	Beam search
Hyperplanes	Likelihood	Branch-and-bound
Naive Bayes	Posterior probability	Continuous optimization
Logistic regression	Information gain	Unconstrained
Decision trees	K-L divergence	Gradient descent
Sets of rules	Cost/Utility	Conjugate gradient
Propositional rules	Margin	Quasi-Newton methods
Logic programs		Constrained
Neural networks		Linear programming
Graphical models		Quadratic programming
Bayesian networks		
Conditional random fields		

(Witten et al., 2016, Ch. 7). Although humans may often have a decent sense of what features might be of relevance, our intuition certainly isn't infallible in this regard, and it diminishes very rapidly with increasingly large feature spaces (imagine trying to track thousands of time-series for individual stock price movements in order to predict the movement of one other stock, for instance) (Domingos, 2012). It therefore can often be helpful to narrow down the candidate models based on the type of task, the size of the dataset, and production requirements (e.g. the computational complexity, storage requirements and costs must be taken into consideration when using ML operationally, as more features and more complex computational tasks can dramatically impact the bottom line) (Witten et al., 2016, Ch. 9).

3.3 A review of electrophysiological signal usage in HA fitting and verification, and considerations regarding the construction of BCI-HAs

3.3.1 Why look at the neural response in this context?

There are a number of reasons that the neural response might be of substantial utility in the fitting process, all of which have been receiving increasing support over the past decade. These include: the fact that central processing mechanisms appear to play an important role in some forms of processing deficits, such as in the speech-in-noise (SIN) deficits that are commonly encountered by HI populations in both aided and unaided conditions (Bidelman, 2017); that age-related declines in important encoding abilities such as temporal synchrony are known to occur across the auditory system, and that AEPs can act as an indicator of

this dysynchrony (Anderson, 2017; Tremblay, Kalstein, Billings, & Souza, 2006); that neural markers of detection and (possibly) of discrimination exist and can be reliably measured in a non-invasive fashion; and that the current standard of audiologic care verifies HAs using real-ear measurements, which give us no information about sound processing beyond the tympanic membrane, and so despite the fact that digital HAs provide audiologists with a great deal of flexibility with respect to the adjustment of frequency settings, these changes in the software of HAs are made without any notion of how they affect the accuracy of the neural encoding of speech (Anderson, 2017).

Although current HA fitting algorithms tend to increase acuity (i.e. audibility), they tend not to increase the clarity of speech sounds (i.e. they result in a distorted signal) (Bidelman, 2017). Since there is evidence to support the notion that AEPs might provide a window into some of the underlying causes of this lack of clarity, it is now believed that an increased understanding of the nature of these neural mechanisms may lead to improved management and assessment strategies (Anderson, 2017; Anderson & Kraus, 2013; Dajani et al., 2013; Easwar et al., 2015a, 2015b).

Generally, it is possible that signal processing performed by HAs may need to be specifically tailored to restore the neurophysiological representation for speech, dependent upon the type of interference that exists in a given listening situation (e.g. noisy vs. reverberant), as well as the specific mix of peripheral and central deficits with which the end-user presents. There is also a prevailing view that in order to accomplish such a restoration, a very detailed understanding of how the normal system works is required (Bidelman, Villafuerte, Moreno, & Alain, 2014). It should also be noted that these neural measures might be used not just in the parameterization of HAs for more optimal performance, but also in the verification of HA fittings, which is especially important in populations where no behavioural measures are available (e.g. infants) or where they are unreliable (e.g. persons with dementia) (Anderson & Kraus, 2013; Hassaan et al., 2016; Munro & Mueller, 2016; Zenker Castro & Barajas de Prat, 2008).

3.3.2 Cortical auditory evoked potentials

The majority of the research aimed at assessing the utility of electrophysiological signals in either determining or validating hearing aid gain settings has been focused on the late or cortical auditory evoked potentials (CAEPs) such as the P1-N1-P2 complex (a.k.a. the acoustic change complex) due to its reflection of acoustic change from silence to sound and from consonant to vowel. To date, indications of the clinical utility of CAEP-related measures remain unclear despite several decades of research (Billings Curtis, Papesh Melissa,

Penman Tina, Baltzell Lucas, & Gallun Frederick, 2012). Their most successful use has been due to their ability to demonstrate audibility with HAs in subjects who are unable to provide behavioural responses (Anderson, 2017; Chang, Dillon, Carter, Van Dun, & Young, 2012), and commercially available technologies such as HEARLab™ are currently being used in clinics for fitting verification. Although there is clear evidence of the utility of the CAEP as a measure of sound detection in both aided and unaided contexts, the evidence of its usefulness in indexing discrimination between sounds is mixed and it should not be used for clinical decision making until a better understanding of the interactions between CAEPs and HA-processed stimuli is attained (Billings Curtis et al., 2012).

Despite the fact that CAEPs are useful in reflecting sound detection abilities, a study by C. J. Billings, Tremblay, Souza, and Binns (2007) showed that changes in presentation level in both aided and unaided conditions in normal hearing adults resulted in amplitude and latency changes, but that equivalent changes in amplification did not. They therefore concluded that 20 dB of gain in the HA was not generally equivalent to 20 dB of presentation level change. The cause for this odd observation is attributed to the fact that the HA appears to be changing more than just the level of the stimulus, and that this leads to some fundamental differences in the way that aided stimuli interact with the central auditory system. In other words, HAs alter the acoustic stimulus, and such processor-related alterations may interact with the way that the auditory system encodes level changes.

One interesting study involving HA users suggests that their CAEPs more closely resemble those of normal hearing subjects while wearing their own personal, previously fitted HAs, which suggest that their brains process speech with greater accuracy and in a more effective manner (Korczak, Kurtzberg, & Stapells, 2005). This was especially true for lower presentation levels. However, most of the change observed in the response is attributed to improvements in the audibility of quiet sounds which is consistent with the general improvements due to HAs (Tremblay, 2015).

A few of different studies have looked at the effects of some mix of presentation level and signal-to-noise ratio (SNR) as they related to CAEPs. One such study by Van Dun, Kania, and Dillon (2016) assessed both normal and impaired subjects using three speech sounds presented at 55, 65, and 75 dB SPL in an unaided condition, and at 55 dB SPL with three different gains in steps of 10 dB. An analysis of the recorded amplitudes revealed no effects of level or gain in normal hearing subjects, but significant effects were found in the HI subjects. Audibility, as determined by signal level or SNR depending on the case, was the dominant predictor of amplitude change. Another study by C. J. Billings, McMillan, Penman, and Gille (2013) looked at how well CAEPs could predict SIN performance in NH adults, and found that SNR was predictive of both CAEP wave characteristics (e.g. latency, amplitude,

and area), as well as behavioural performance, but that sound level was only predictive of behavioural performance. Generally, studies suggest that SNR is a key contributor to CAEP morphology, and that it is particularly important when establishing the effects of HAs on auditory processing (Billings Curtis et al., 2012) especially when examining aided CAEPs for assumed level effects, as any changes (or lack thereof) might be due to audible amplified noise (i.e. SNR cues) rather than signal level cues (C. J. Billings et al., 2013).

3.3.3 Auditory steady state responses

The auditory steady-state response (ASSR) has also been explored with respect to its utility in the fitting and validation of HAs, although not to the same extent as CAEPs. Much of the research has been along the same lines, however, as the ASSR has primarily been investigated for its ability to provide objective threshold estimates, the prescription of gain in hard-to-treat populations such as infants, and in the verification of amplification through detection for the sake of adjustment when necessary (Sardari, Jafari, Haghani, & Talebi, 2015; Zenker Castro & Barajas de Prat, 2008). Interestingly, there have been some positive results indicating the potential use of the ASSR as a correlate of loudness growth in both normal and impaired hearing populations (Emara & Kolkaila, 2010; Van Eeckhoutte, Wouters, & Francart, 2016; Zenker Castro & Barajas de Prat, 2008; Zenker Castro, Juan Barajas de Prat, & Larumbe Zabala, 2008). There is also some recent evidence indicating that there is less difference between evaluations of functional (i.e. behavioural) gain and ASSR gains as assessed for HA validation than there is between behavioural and ASSR threshold data, and thus comparing gains rather than thresholds might be more accurate for the validation of HAs (Sardari et al., 2015).

This is all very promising, in part because it supports the possibility of creating and fitting more individually tailored, objective, and automatically fitted HAs in the future, but also because the ASSR is itself a frequency following response in a literal sense (Picton, 2010), and there is good reason to believe that its underlying generators are largely the same generators as those responsible for the steady-state portion of the FFR, although the cortical contributions may vary slightly depending on the frequency of the modulation (Herdman et al., 2002). Assuming this is the case, the difference lies in the stimuli employed, with the typical ASSR stimulus being comprised of one or more narrowband carrier frequencies each modulated at a given frequency (often 40 or 80 Hz), as well as the typical testing paradigm employed, which in the case of the ASSR usually comprises simple detection of the signal in the frequency domain with respect to its amplitude and phase. Thus, the major difference between typical ASSR recording and the FFR recording in the two recent studies by Easwar

et al., for instance, is the more ecologically valid, behaviourally relevant broadband stimulus employed to elicit the FFR. Of course, it is extremely important to note that at lower modulation rates (e.g. 40 Hz), which are not typical of the speech stimuli employed in FFR studies, the ASSR is susceptible to effects of arousal and vigilance, and is known to be appreciably altered by sleep, for example, whereas the FFR is not.

3.3.4 Auditory brainstem responses to brief stimuli

The auditory brainstem response (ABR) to brief stimuli has been studied fairly extensively for a variety of reasons, including for its utility as a measure of the general integrity of the peripheral and central auditory system in newborns, to detect tumours and lesions, and to determine the type and degree of hearing loss in impaired subjects (Burkard et al., 2007; Picton, 2010). It has also been used to estimate behavioural thresholds, and a variety of correction factors have been developed to achieve this, both for the click, chirp and tone-burst ABRs (Hall, 2013; D. Stapells et al., 1996; D. R. Stapells, 2000; Xu, Cheng, & Yao, 2014). HA fitting formulas, such as the DSL v5.0 (Scollie et al., 2005) have been devised to make use of these thresholds, and are used in difficult-to-treat populations.

Under the assumption that ABR amplitude is directly related to loudness perception and not merely to behavioural thresholds (Da Silva, 2009; Korczak, Sherlock, Hawley, & Formby, 2017; Serpanos, O'malley, & Gravel, 1997; Wilson & Stelmack, 1982), the idea that HA gain profiles could be set to achieve ABR amplitudes characteristic of various loudness levels has been put forth (Da Silva, 2009; Hoseingholizade, 2015). Although this approach cannot be discounted, the general view is that ABRs to brief stimuli are not likely to yield as much useful information as their alternatives in the fitting context, primarily due to the fact that they tend to be much less frequency specific (especially in the case of the click), and the duration of the stimulus presentation is not very compatible with the time constants of most HAs, and so they are not useful in understanding the effects of compression for instance (Anderson, 2017; Hassaan, 2011). Furthermore, the stimuli employed are not as behaviourally relevant as speech tokens.

3.3.5 Frequency following responses

Relatively little has been done to investigate the utility of the FFR in comparison to the above-mentioned AEP measures. This is surely related to the increased complexity typical of the FFR recording setup in comparison to the ABR and the ASSR, both with respect to the speech stimuli typically employed, as well as to the increased recording time needed to create meaningful averages with the characteristic temporal and spectral features that are

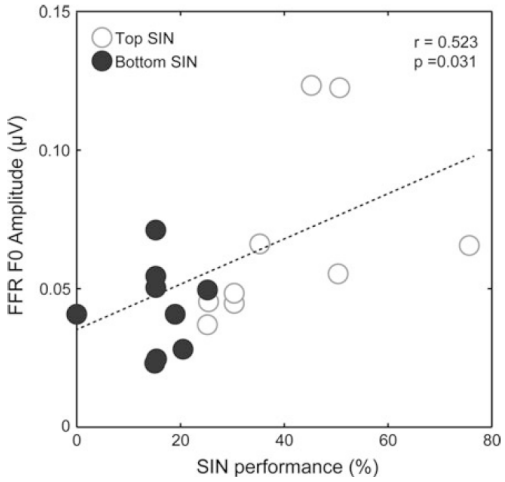

Figure 3.3: SIN performance and the FFR . Correlations between normal hearing subjects in the top and bottom quartiles of behavioural speech-in-noise performance and their frequency following response amplitude at the fundamental frequency. Larger amplitudes are correlated with better performance. Reprinted with permission from "Communicating in Challenging Environments: Noise and Reverberation", p. 209, by G. Bidelman. In The Frequency Following Response, by N. Kraus et al. (Eds.), 2017, Switzerland: Springer Nature. Copyright 2017 by Springer International Publishing AG.

the focus of most subsequent analyses (Bellier et al., 2015). It is also due to the much smaller amplitude of the FFR in comparison to the cortical potentials, which are discernible even to the untrained eye from presentation to presentation (see Fig. 2.1) (Picton, 2010). Recent evidence supporting the sensitivity of the FFR to a slew of relevant auditory processing features and pathologies (see Kraus, Anderson, White-Schwoch, Fay, and Popper (2017) for a recent review), and in particular those that are most problematic for HA users (e.g. a lack of clarity and abnormal loudness) has led to a surge in interest.

Fortunately, recent research both from our lab and elsewhere indicates that the FFR provides a sensitive measure of both of the key disruptors that contribute to a lack of clarity, namely noise and reverberation (Al Osman, 2016; Al Osman, Giguère, & Dajani, 2016, 2018; Bidelman, 2017; Laroche, 2010; Laroche, Dajani, Prévost, & Marcoux, 2013; Prévost, Laroche, Marcoux, & Dajani, 2013). Particularly of interest with respect to HA optimization is the fact that the FFR appears to be predictive of SIN performance (see Fig. 3.3), and that it seems to provide an index of physiological (or internal) SNR (Dajani et al., 2013; Prévost et al., 2013). Since these noisy and reverberant conditions are especially difficult for HA users, an objective measure of their neural impacts is crucial in understanding the effects of amplification and how it might be adjusted to improve the overall performance of HAs in such environments.

One of the key advantages of the FFR over other AEP measures is that it provides a window into the neural processing of both the envelope and temporal fine structure of speech sounds (Bidelman, 2017; Bidelman & Howell, 2016). Given the prevailing view that normal auditory function is reliant on cues extracted from both of these, and that these

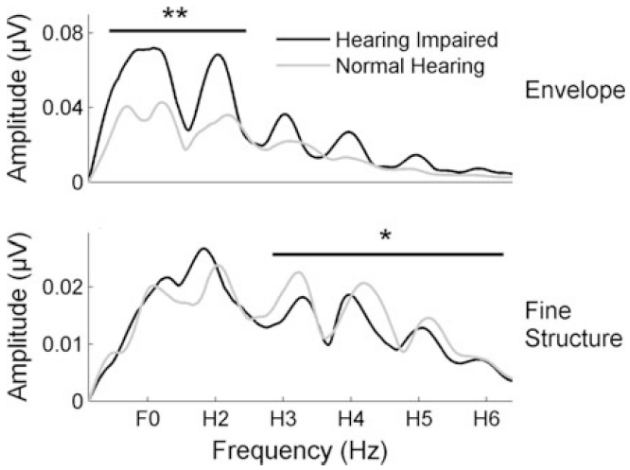

Figure 3.4: ENV and TFS differences between NH and HI populations. Clear differences between normal (gray) and impaired (black) hearing responses to the (C) envelope and (D) temporal fine structure can be observed. Here, the hearing impaired subjects demonstrate larger envelope components at the fundamental and its nearest harmonics, whereas the normal hearing subjects exhibited higher amplitudes in their encoding of the higher harmonics of the temporal fine structure. Reprinted with permission from "Clinical Translation: Aging, Hearing Loss, and Amplification", p. 281, by S. Anderson. In The Frequency Following Response, by N. Kraus et al. (Eds.), 2017, Switzerland: Springer Nature. Copyright 2017 by Springer International Publishing AG.

cues are exploited to varying degrees by different hearing populations under various conditions (Anderson, 2017), the ability to understand just how these are being impacted at the neural level is critical to a forward-looking strategic approach to HA optimization. One key example here is that there is evidence to suggest that older adults utilize TFS cues for speech comprehension in reverberant conditions, whereas younger adults use ENV cues, and so speech intelligibility suffers more in older adults in reverberant conditions with additional noise since TFS is known to be more disrupted by noise (Bidelman, 2017).

Further to this, Anderson et al. (2013) and Anderson and Kraus (2013) showed that there appear to be important differences between the eFFR and sFFR of NH and HI populations, as the envelope features observed actually tended to be larger in HI individuals (see Fig. 3.4). This result is consistent with other studies, and it has been conjectured that the enhancement of envelope cues might actually distract from the fine details required for adequate speech comprehension. This supports the notion that there is a need to balance the contributions of ENV and TFS, and that this might be done through the windows of the eFFR and sFFR. Although these results are promising, much more research is needed here as there are discrepancies between some of the reported results (Ananthakrishnan et al., 2016; Anderson et al., 2013), and the balance in the encoding of ENV and TFS is likely to depend on a variety of factors, including the specifics of the subject's hearing loss, as well as the frequency content, bandwidth, and presentation level of the stimulus.

Although several of the tabulated studies (Table 2.3) confirm that the FFR can be recorded in an aided condition, and that there exist significant contrasts between NH and

HI populations in both aided and unaided conditions, altogether reinforcing the need to further understand the impacts of hearing loss and amplification on ENV and TFS, only two studies actually explored the differential impacts of two separate amplification parameterizations. The first report on the potential utility of the FFR to objectively determine the appropriateness of HA fitting is due to a case study on a single subject (Anderson & Kraus, 2013; Kraus & Anderson, 2012). Responses to the two different HA settings were recorded via free field presentation, and the setting that yielded the highest correlation with the stimulus was indicated to be the preferred setting of the subject (Kraus & Anderson, 2012), which resulted in the authors putting the measure forth as a potential metric for performance verification.

Unfortunately, very few details are provided, and it is not clear whether the subject is hearing impaired, whether or not they have worn hearing aids and if so for how long, how old they are, just how the correlation between response and stimulus was performed, and what the specifics of the two HA settings are. Further to this, there is reason to suspect that such a simple correlative metric might not be the best indicator of the quality of the internal representation of sounds due to the fact that the FFR is known to reflect composite, nonlinear transformations in the auditory periphery and in the brainstem (Dajani et al., 2013).

The second study by Bellier et al. (2015) provides a much more detailed assessment of the feasibility of recording FFRs through direct, wireless HA stimulation which reduces jitter due to head movements. The results of the four NH subjects recorded indicate that artifact-free FFRs are indeed feasible under such a setup and that significant alterations in the spectral and envelope FFRs exist between settings as shown in Fig. 3.5. Although they did not study different HA settings, Easwar et al. (2015b) did further validate the ability to record FFRs in aided HI subjects, and they importantly showed that FFR amplitude and detection measures are positively correlated with speech discrimination scores and sound quality rating, suggesting that they might be useful for predicting suprathreshold hearing performance.

3.3.6 On the validity of using vowels and a machine learning approach to simulating categorical perception

Work from our lab has already established that machine learning can be used to accurately predict vowel categorization based upon information present in the FFR (Sadeghian, 2012; Sadeghian et al., 2011, 2015). Two additional studies that assess vowel classification using FFRs have also been published (Won et al., 2016; Yi et al., 2017), both of which use formant

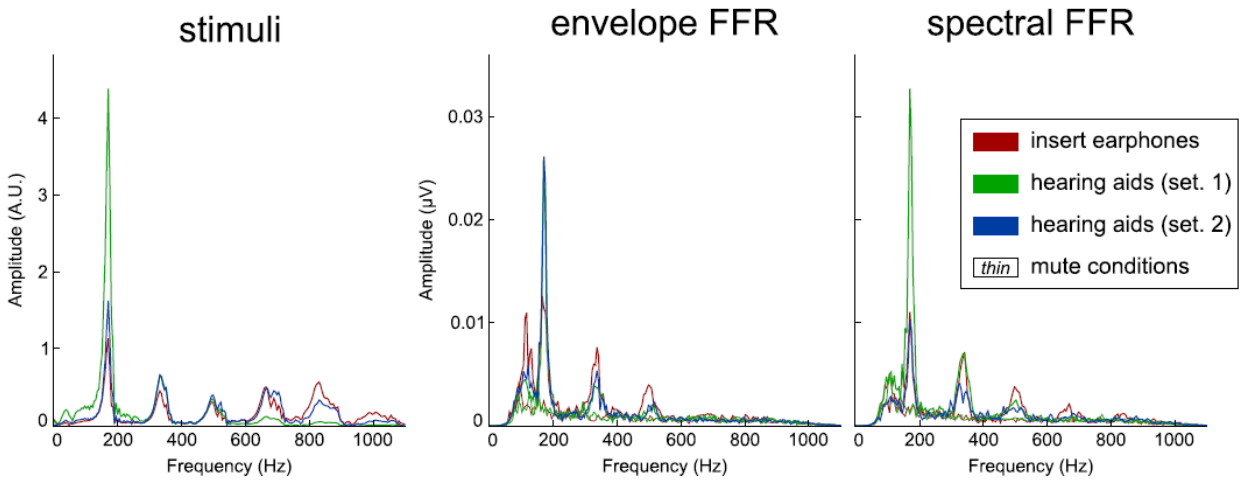

Figure 3.5: FFR spectra under different hearing aid settings. Spectra of /a/ vowel as presented through various transducers indicated per the inset legend, and their corresponding envelope and spectral FFRs. Reprinted with permission from "Speech Auditory Brainstem Response through hearing aid stimulation", by L. Bellier et al., 2015, Hearing Research, 325, p. 53. Copyright 2015 Elsevier B.V.

features as inputs in order to maximize classification accuracy in their selected models. Another study by Jafarpisheh et al. (2016) examined the potential utility of a nonlinear feature extraction process for the objective classification of /da/, /ga/, and /ba/ syllables, concluding that the extracted feature consistently indexes the formant transition period which is known to contain information that differentiates between these consonant-vowel classes. It seems no small coincidence that the nonlinear categorization function that Bidelman, Moreno, and Alain (2013) recorded in the cortical evoked responses to vowels - and that they subsequently modeled as a cortical transform on the first formant frequency as represented by the FFR with striking agreement - turns out to be sigmoidal in nature. Sigmoidal activation functions are commonly used in machine learning (e.g. perceptrons) (Uykan & Koivo, 2000), since they can take continuous inputs (e.g. F1 frequency) and transform them to bound ouputs between zero and one, which can then be interpreted as the probability of belonging to a given category or not. They are also very computationally cheap.

This leads to the question of whether or not machine learning can be used to effectively determine whether one amplification setting might lead to more or less categorization errors based on FFR inputs to the selected algorithm alone. That is, can an ML algorithm be selected that reliably behaves similarly to the categorization mechanisms of the cortex, and can we therefore simulate HA impacts on behaviour using the FFR alone with a good degree of accuracy? And can this be done in a way that is invariant to sound level, as is largely accomplished by the auditory system? And further to this, can the sound level be accurately predicted from the FFR? If the answer to these questions is yes, then it stands to reason that the BCIs built to exploit the FFR for HA optimization purposes would be able to

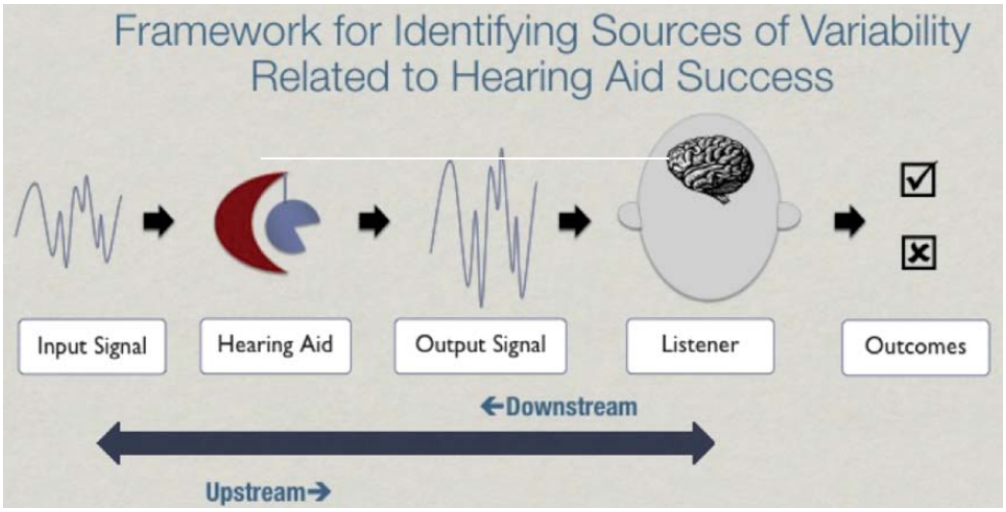

Figure 3.6: Framework for identifying sources of variability related to hearing aid success. Tremblay's (2015) schematic depiction of the components that contribute to the variability related to the success of hearing aid outcomes. Note the explicit acknowledgement of the brain as a source of variability and a contributor to both upstream and downstream impacts. A variety of auditory evoked potentials allow for the capture of some of the neural variability related to such successful outcomes, and should therefore be considered in any end-to-end system aimed at capturing and exploiting this variability via tailored and targeted rehabilitative strategies. Reprinted with permission from "The Ear-Brain Connection: Older Ears and Older Brains", by K. Tremblay, 2015, American Journal of Audiology, 24, p. 118. Copyright 2015 by American Speech-Language-Hearing Association.

control, to some degree at least, for the common issues of loudness and lack of clarity so often encountered. Incorporating such a BCI into SFHAs, as discussed in Section 2.3, may eventually help to remove the barriers to accessing appropriate amplification that many hearing impaired persons currently experience.

Ultimately, the literature supports the notion that AEPs can generally be exploited – and likely with mutual benefit in an integrated system – in order to help to identify sources of variability related to hearing aid success, as well as to devise and validate new therapeutic solutions, including those that are centred around amplification. A deep understanding of all of the components outlined in Tremblay's schematic for such hearing aid success is on the horizon (see Fig. 3.6). With rapid advances in technology, our understanding of auditory system components, and in the development of machine learning tools, it is likely that we will be able to very accurately simulate both upstream and downstream interactions between components from end-to-end across the chain in order to more successfully predict outcomes in the decades to come, and to therefore disburden – to the extent that is biologically possible – those who suffer from hearing impairment.

3.4 Gap analysis: towards a better understanding of the FFRs utility in a BCI-HA

The background information and review of the literature reveals the following:

1. Methods for the successful recording of aided FFRs have been developed, and studies focusing on HI populations are becoming increasingly common. To-date, no studies providing a thorough analysis of both the eFFR and sFFR in the time and frequency domains has been published.

2. Age is often a significant confounding variable in typical HI / HA studies, and should be accounted for in FFR-based study designs assessing BCI-HA constructions.

3. There is a need for more behaviourally relevant stimuli in FFR studies in general, but especially as pertains to HA-related studies owing to the well-known existence of distortions related to compression time constants and other HA features such as feedback reduction settings. Brief phrases or sentences that can be shown to elicit the effects of such features should be utilized in order to determine the real-world behaviour of the HA via the FFR.

4. The FFR appears to be a robust indicator of the 'preprocessed' signal as it exists prior to 'feature extraction' and 'classification' that appears to occur in the thalamic and cortical areas. This makes it a solid candidate target for exploitation with respect to BCI-HA construction.

Given the recent establishment of feasibility of recording aided FFRs, it is only a matter of time before a variety of effects are studied on HI populations in an aided condition. This will include not only amplification, but investigations of other important parameters and settings such as compression time constants and ratios, noise, wind and feedback reduction settings, the benefits of directional microphones, etc. (Dajani et al., 2013). In order to provide context to these results, and to understand how HA adjustments might be made in order to facilitate more optimal performance, it will also be necessary to understand the response of NH subjects under similar conditions. Several caveats exist here however, and the utility of the NH model as a study tool for HI in general appears limited to some extant due to the fact that it does not appear we can fully restore BM mechanics and AN firing with via HAs (Giguère, 1998; Giguère & Smoorenburg, 1999). Therefore, it is unlikely that even an optimal, custom-made HA (e.g. with specific frequency channels to account for cochlear dead regions, perhaps) would enable the total restoration of normal peripheral auditory processing.

Another important issue in the understanding of HI also arises here, in that the effects of increasing age and pathophysiology are often confounding factors that are not easily disentangled since they tend to co-occur (Anderson, 2017; Anderson et al., 2013). Fortunately, it is possible to use the FFR to study both of these factors and a further elucidation of their various impacts attributable to normal aging and SNHL due to insult or pathology is likely to guide the rehabilitative strategies of HI. These strategies might include training where temporal encoding is diminished by the effects of aging, for instance, or amplification where the degradation is not attributed to normal age-related declines in synchrony but is indicative of SNHL (Anderson, 2017).

Ideally, a deeper understanding of both the effects of sound level and of HA settings and their interactions will be realized across not only the entire phonemic catalogue, but also across the space of common words and sentences in the native language and dialect of the end-user (Chasin, 2011). This is important because inappropriate loudness due to amplification is believed to negatively impact the clarity of speech among HA users. Here, it would be ideal to understand how to control for loudness while simultaneously increasing the clarity experienced by the HA user. The use of more realistic word and sentence-length stimuli should also not be overlooked (C. J. Billings, Penman, McMillan, & Ellis, 2015; Koerner, Zhang, Nelson, Wang, & Zou, 2016; Reichenbach, Braiman, Schiff, Hudspeth, & Reichenbach, 2016), since it might shed light on the importance of compression time constants, for instance, and there is evidence to suggest that auditory system appears to make use of a dual-scale temporal processing scheme wherein a local scale uses short temporal windows to extract fine-detail acoustic information, and a global scale uses long temporal windows to abstract acoustic patterns (Teng, Tian, & Poeppel, 2016). A phoneme-level approach, while extremely valuable as a first approximation to our understanding of neural processing, is likely only to be valuable in that it serves as a touchstone for further, more complex analyses that are more representative of the auditory environment humans typically find themselves in.

Such sentence-length stimuli - as used by Reichenbach et al. (2016), for instance - would also allow for a deeper understanding of the trade-offs that appear to exist between the complex impacts of cognitive function, neural plasticity, pathophysiology, and the altered stimuli that result from HA parameterizations (C. J. Billings et al., 2013; Korczak et al., 2005; Song, Skoe, Banai, & Kraus, 2011). As an example here, recent evidence suggests that when audibility is controlled for, the difference in behavioural SIN performance is likely attributable to differences in the activation of higher cortical centres (i.e. those beyond the primary AC) that are associated with language processing (Bidelman & Howell, 2016; C. J. Billings et al., 2015; Williams-Sanchez, 2014).

Despite the need for more behaviourally relevant stimuli in general, a strong case can be

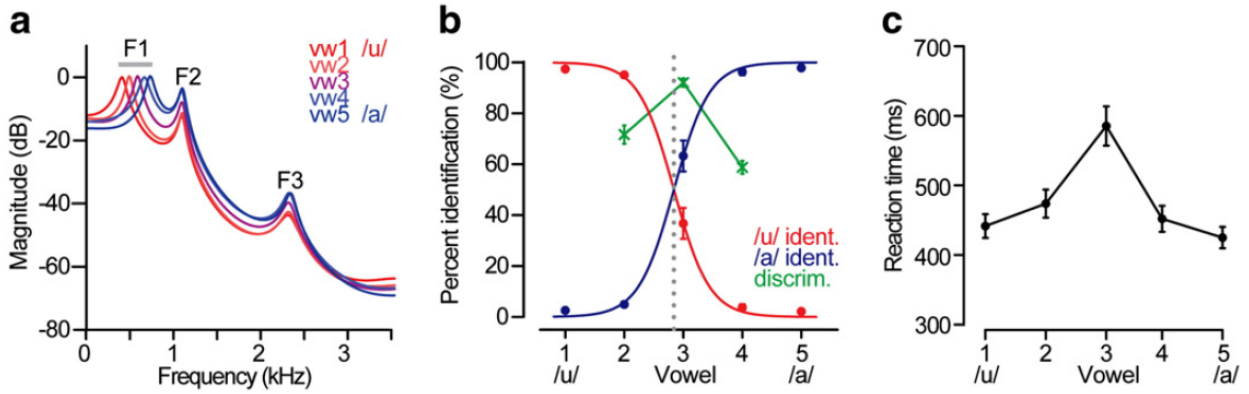

Figure 3.7: Speech vowel continuum used to probe categorical speech perception via auditory evoked potentials. A – Spectral envelopes of the five vowel stimuli along the /u/ to /a/ continuum, as determined by first formant (F1) location. B – Psychometric identification and discrimination functions, with the categorical boundary denoted by the dashed line. C – Reaction times for vowel identification. Although the change in F1 is continuous in frequency, subjects hear a discrete perceptual shift in phonetic category around the boundary, and are much slower at labeling vowel 3 as compared to within boundary tokens (e.g. 1-2, and 4-5). Reprinted with permission from "Tracing the emergence of categorical speech perception in the human auditory system", by G. Bidelman et al., 2013, NeuroImage, 79, p. 203. Copyright 2013 by Elsevier Inc.

made for the use of vowels at present, since they not only represent a step in the right direction in this regard when compared to the stimuli more typical of the present-day clinicians' AEP toolbox, but also due to the fact that there is evidence that the spectral proximity of the first formants (F1) often results in confusion between stimulus pairs in noise (Nabelek & Letowski, 1985), that F1 encoding in the FFR is a significant indicator of vowel recognition performance (Won et al., 2016), and that F1 separation is critical in stream segregation (Alain, Arsenault, Garami, Bidelman, & Snyder, 2017), which is a particularly difficult task for HI individuals in noisy environments. Further supporting the use of vowels is the fact that Bidelman et al. (2013) demonstrated that cortical responses to vowel stimuli with F1 near the categorical boundary corresponded with the categorical perceptual assessment of the listeners, despite the lack of apparent differences in the encoding of the corresponding FFRs (see Figs. 3.7 & 3.8). Not only does this support the prevailing view that categorical perception emerges at the cortical level, it also supports the view that the FFR may not reflect a true neural correlate of the auditory percept elicited by a stimulus, but rather that it may reflect more exogenous stimulus properties (Bidelman, 2017; Bidelman et al., 2013; Gockel, Carlyon, Mehta, & Plack, 2011).

That the FFR appears to be more closely related to stimulus properties than it does to perceptual qualities is compatible in some sense with the view that the brainstem and midbrain stages of processing are not dissimilar to the 'preprocessing' stage of the typical machine learning workflow presented in Fig. 3.1. Since the medial geniculate body appears to transform the still highly synchronous response present in the inferior colliculus to a rate code as it transmits information to the primary auditory cortex, it stands to reason that

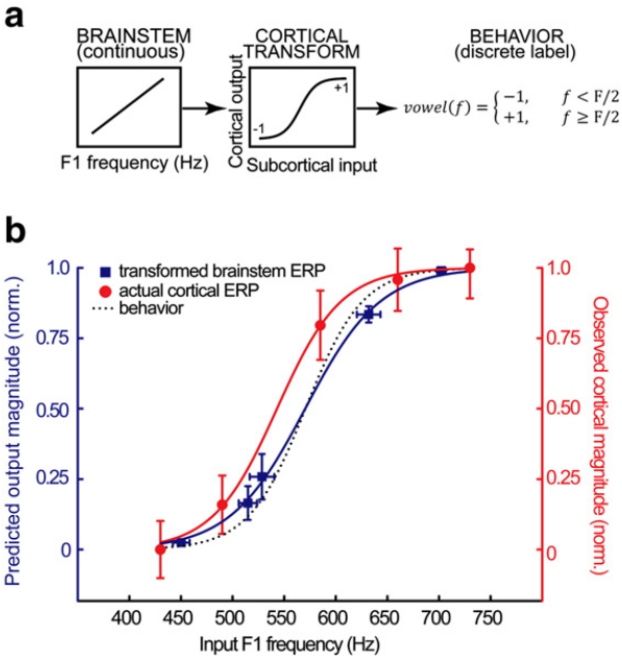

Figure 3.8: Acoustic phonetic transformation in neuronal speech representation. A – Schematic depicting how a continuous acoustic paramenter such as F1 gets mapped from its faithful representation in the brainstem into adiscrete perceptual code in the cortex via a nonlinear, sigmoidal transform. B –Agreement between behavioural responses, cortical responses, and predicted cortical responses based on transformations of brainstem responses per the model in A. Reprinted with permission from "Tracing the emergence of categorical speech perception in the human auditory system", by G. Bidelman et al., 2013, NeuroImage, 79, p. 208. Copyright 2013 by Elsevier Inc.

these two layers might then correspond to the 'feature extraction' and 'classification' layers, respectively. Although this is merely hypothetical in nature at present, relatively simple and biologically plausible feedforward neural network models trained using the so-called infomax principle, which maximizes the transmission of information from input to output (and so too to the intermediate layers as a result), have been shown to develop receptive field properties similar to those of auditory nerve, cochlear nucleus, medial geniculate body, and auditory cortex neurons (Tanaka, 2016).

This same principle has also been shown to result in networks with layers that have similar receptive field properties as the visual cortex neurons involved in well-known tasks such as edge detection (Tanaka & Nakamura, 2013), and could therefore point towards a common and extremely plausible principle underlying the cost-function of the brain's neural networks: namely, the minimization of information loss. This is in line with recent insights from the fields of machine learning and artificial intelligence, which suggest that deep learning networks organized in configurations similar to those known to exist in the brain result in sparse coding or sparse representation of their input space (G. Billings, Piasini, Lőrincz, Nusser, & Silver, 2014; Schmidhuber, 2015; D. Wang, Zhou, & Hussain, 2016), with output layer neurons being highly selective (i.e. very narrowly tuned) and seemingly matching the

probabilistically salient features of natural stimulation (G. Billings et al., 2014; Vidya, Nasira, & Priyankka, 2014). This behaviour is precisely what would be expected from a neural code that conforms to the information theoretic principles of not only the maximization of information transmission, but also the minimization of description length (G. Billings et al., 2014; Foldiak, 2003). The latter effectively amounts to compression, and as a whole this strategy minimizes the amount of energy expended by the system (e.g. the brain), since the total number of spikes is inversely proportional to the sparseness of the representation (G. Billings et al., 2014; Foldiak, 2003; Schmidhuber, 2015).

Chapter 4
On the FFR representation of vowels presented at different sound levels

4.1 Introduction

The first two studies aimed to answer the following questions, respectively:

1. How does the representation of a long vowel change with level in the FFR of normal hearing subjects?

2. How does the representation of four short vowels change with level in the FFR of normal hearing subjects? Do these representations differ between sexes? Are they stable across test and retest conditions?

Chapter 3 provided a detailed overview of the relevant background information and the underlying reasons for performing these two studies. Briefly, the objectives were to not only provide a more thorough characterization of the level-related changes in the FFR of normal hearing adults with respect to the encoding of envelope and temporal fine structure cues in both the time and frequency domains, but also to create a dataset for future use in a machine learning study (see Chapter 5.5) which could in turn inform an assessment of the feasibility of the BCI-HA technology space (see Chapter 6).

To these ends, we first sought to record the response of normal hearing adults when presented with a fairly realistic stimulus – in terms of frequency content, level and duration – and to then compare these results with the results of a broader set of stimuli that are otherwise very similar in their construction aside from their more abbreviated duration. Here, the rationale with respect to the ultimate research goal was that these results would provide insight into the encoding of level across the conversational speech range and into how duration might impact an automated fitting process. Since it was not generally known what type of stimuli might best serve such a fitting process, English vowel sounds were selected to constitute the stimulus set since they are fairly primitive yet meaningful speech units and they have been studied quite extensively using similar methodological considerations and analytical tools in the literature (see Sections 3.3.6 and 3.4).

In an ideal world, recordings of all of the relevant consonants and vowels across the entire reasonable range of level, duration, frequency space, pitch contour, context etc. and their

various combinations would be possible, and this would permit a rigorous assessment of what a minimally sufficient stimulus set construction would entail in the proposed fitting context. In order to accomplish this, exhaustive tests of hearing aid users would need to be performed and some kind of tree graph-based algorithm would ultimately have to be developed to select an optimal stimulus set depending on each user's hearing profile and the programming options of their device. Obviously, given the relatively long duration required to record the FFR to one speech stimulus, one can ill-afford any attempt to come close to such a complete characterization of the space, and so the stimuli here were chosen by making thoughtful compromises along the following lines:

- the stimuli should be comparable to others used in the literature where this does not in any obvious way diminish the envisioned explanatory power of the results in order to invite comparisons to what is already known
- the mutual information content of the stimuli and the recordings should be substantial
- the stimuli should be created so as to make the duration of both the individual recording sessions and the research project as a whole realistic, and in accordance with the ethics approval for the project
- the stimuli should in some meaningful sense span the space of speech sounds so as to constitute some form of sensible covering in order to maximize the possibility of assessing the feasibility of the approach

With respect to the second consideration, both the nature of the FFR as a signal as well as the abilities of the recording equipment came into play. Previous results, as well as the results from our pilot studies indicated that higher frequency components of the stimuli are often not discernible using the recording equipment available in our lab. Although some newer systems have emerged in the last decade that might permit a better characterization of the high frequency components of the FFR, our equipment has the limitation of having a fixed maximum number of samples regardless of the duration of the stimulus and calculates the averaged FFR online. Additional limitations related to the collection equipment and software constrains the stimulus selection process (e.g. a discrete set of recording window lengths), but not in any way that was found to be excessively compromising.

Taking all of these considerations into account led to the selection of the five different vowel stimuli – a 300 ms /a/ vowel for Study 1, and 100 ms /a/, /ɔ/, /U/, and /u/ vowels for Study 2 (see Table 4.1 for details). Not only were the selected stimuli extremely similar to those previously used in our lab, they also provided substantial overlap with the relevant literature, especially with the two studies by Krishnan (1999, 2002) and the study by Akhoun

et al. (2008) performed on normal hearing adults (see Table 2.3). This was done in order to permit fairly direct comparisons with previous results and to provide some adequate covering of the vowel space within the limits of the frequency range typical of the FFR.

It was hypothesized that in both experiments some of the features known to be of relevance to speech would be robustly encoded for all stimuli within the phase-locking limits typical of the brainstem (e.g. F0, F1, and perhaps F2), and that differential effects of increasing level would be observed primarily through general increases in the spectral amplitude of F0 and its harmonics and via decreases in latency of waveform components. By testing these hypotheses, we aimed to contribute to the limited body of work that has addressed such level-related changes in the FFR (described in Section 2.3).

In addition to the above, determining whether or not the speech FFR is stable both within a testing session and from one testing session to another is of significant importance. It is of interest from a pure research perspective not only because the test-retest reliability of the speech FFR has not yet been examined at multiple levels and across multiple stimuli to our knowledge, but also due to the fact that any lack of test-retest reliability here might imply that there are top-down effects modulating the response and this could possibly be due to some interaction with sound level. From a practical or clinical perspective, any substantial variability here would be cause for concern as the diagnostic or prognostic utility of any derivative measures would be compromised. And within the context of this work, high test-retest reliability among responses would of course have highly positive impacts on any related ML or BCI-HA endeavors.

It is worth emphasizing here that test-retest stability is especially required for vowel durations that are typical of conversational speech. Work by Van Santen (1992) showed that stressed and unstressed vowels typically had durations of roughly 80 ms and 130 ms, respectively. Therefore, the utilization of 100 ms vowels in Study 2 seems appropriate for such an analysis.

Finally, since there are known sex differences with respect to certain AEPs in terms of both latency and amplitude (see Section 2.4), and since the results of Study 1 seemed to indicate potential level-related differences, Study 2 sought to determine whether or not any such sex-level interactions were occurring. Meaningful differences here might also impact ML and BCI strategies as they might require sex-specific solutions. Imagine here that latency is a key predictor of level, but that systematic differences in latency exist between male and female populations. Training an ML model (absent a categorical variable for sex) on data from both sexes would likely result in many level confusions as male responses would be predicted to be due to lower-than-actual sound levels, and vice-versa for females.

4.2 Methods

The experimental designs and undertakings for these two studies were nearly identical, and they will therefore be treated here under the same umbrella, and should be considered to be identical whenever there is no explicit mention of their divergence.

4.2.1 Subjects

Fourteen normal-hearing English speaking adult subjects (6 female) ranging from 19-46 years of age participated in the first study (NB: only one subject was in their 40's, and the next oldest subject was 36 years old). Prior to recording, each subject underwent an audiometric test to verify that they possessed hearing thresholds of 20 dB HL or less at 250, 500, 750, 1000, 2000, 4000 Hz in both ears (using a pure-tone air-conduction test). Participants were seated comfortably in a reclining chair inside of a sound attenuating booth that conforms to ANSI standards for background noise level (ANSI S3.1-1999 (R2013)) and were encouraged to stay awake and as still as possible throughout the duration of the recording session. All subjects provided informed consent and all recordings were performed in line with the requirements of the University of Ottawa's Ethics Board (approval notice file number H06-09-05, see Appendix I). Two subjects (1 female) were excluded due to excessively high impedances ($>$ 6 kOhms).

Study 2 is identical, but a total of 22 English speaking adult subjects (11 female) aged 20-36 years of age were recorded over two sessions (a minimum of 3 days apart, and occurring 23 days apart on average), with the audiometric testing being performed prior to the commencement of their first recording session. Here, the subjects were permitted to watch a muted, subtitled documentary since it was observed that some subjects from Study 1 were having difficulty remaining alert throughout the duration of the recording session. Two additional female subjects were recruited, but one failed the audiometric threshold testing and the other divulged during the first session that she had very recently been referred to an ENT by an audiologist despite having normal audiometric testing, and was excluded as a result.

4.2.2 Stimulus creation and calibration

As mentioned, Study 1 made use of a 300 ms /a/ (as in 'father) vowel sound and Study 2 made use of four 100 ms vowel sounds: /a/ (as in 'father'), /ɔ/ (as in 'ball'), /U/ (as in 'hood'), and /u/ (as in 'who'd). All stimuli had a common F0 of 100 Hz, and the two durations were selected in order to have an integer number of periods at F0. The stimuli were

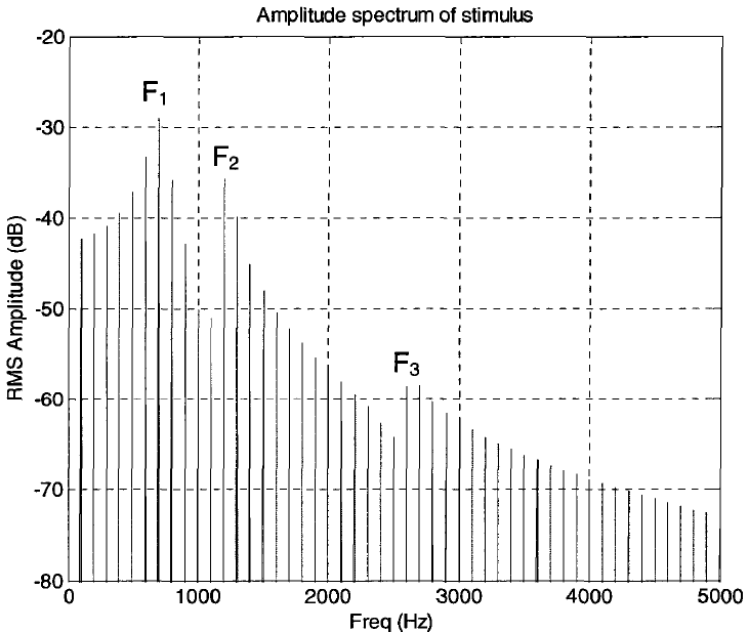

Figure 4.1: Spectrum of /a/ vowel with the first three formants, indicated as F1, F2, F3, respectively. The fundamental frequency is 100 Hz, and so there is power at the fundamental and its harmonics. Reprinted with permission from A Study of Auditory Speech Processing using Brainstem Evoked Responses under Quiet and Noisy Conditions (p. 11), by M. Laroche, 2010, Ottawa, Canada: University of Ottawa. Copyright 2010 by Marilyn Laroche.

Table 4.1: Stimulus parameters for vowel duration, fundamental frequency, formant frequencies, bandwidth, and relative levels. The 300 ms /a/ vowel was used in Study 1, and the other four stimuli were used in Study 2.

Vowel	**Dur. (ms)**	**F0 (Hz)**	**F1 (Hz)**	**F2 (Hz)**	**F3 (Hz)**	**BW1 (Hz)**	**BW2 (Hz)**	**BW3 (Hz)**	**A1 (dB)**	**A2 (dB)**	**A3 (dB)**
/a/	300	100	700	1200	2600	130	70	160	-1	-5	-28
/a/	100	100	700	1200	2600	130	70	160	-1	-5	-28
/ɔ/	100	100	600	900	2400	100	60	110	0	-7	-34
/U/	100	100	500	1200	2200	80	100	80	-1	-12	-34
/u/	100	100	300	900	2200	65	110	140	-3	-19	-43

created using a simplified Klatt formant synthesizer at a sampling rate of 48 kHz and at 16-bit resolution (consistent with its previous usage in our lab – see Laroche (2010); Sadeghian (2012)) and were synthesized according to the parameters indicated in Table 4.1 (see Fig. 4.1 for an example of the stimulus spectra). The values in the table represent slight modifications to the average formant frequencies provided in the classic study of vowels by Peterson and Barney (1952) while staying within the formant space boundaries representative of each vowel (see Appendix A.3). This was done in order to ease the extraction of formant energy.

All stimuli were presented to each subject at four levels ranging from 55 to 85 dBA in 10 dB increments. The stimulus levels were calibrated with the insert earphone connected to a 2-cc coupler (Bruel & Kjaer DB0138 with microphone Bruel & Kjaer Type 4144), and subsequently to a Bruel & Kjaer Sound Level Meter Type 2235.

4.2.3 Experimental setup

The FFRs were recorded using the Bio-logic BioMARK v.7.0.2 system (Biological Marker of Auditory Processing, Bio-logic Systems Corp, Mundelein, IL; note that the Bio-logic product line is now distributed by Natus Medical Inc., Pleasanton, CA). The stimulus was delivered to the right ear through a Bio-logic adult-sized foam-tip insert earphone, and the left ear was kept unconcluded throughout the duration of the experiment. All of the recordings were made with non-disposable Bio-logic silver/silver chloride electrodes, with the active electrode placed at vertex (Cz), and the reference and ground electrodes placed on the right and left earlobes, respectively. In order to obtain impedances of less than 6 kΩ, a mild abrasive and a conductive EEG paste (NuPrep and Ten 20 Conductive EEG Paste, respectively – both by Weaver and Company, Aurora, CO) were used on the skin of the participants. The acquired EEG signals were amplified with a gain of 100000, and were bandpass filtered online between 30-1000 Hz in the first study, and 100-3000 Hz in the second. An artifact rejection criterion of 23.8 microvolts was applied to suppress myogenic contamination. The recordings were synchronized to the stimulus onset, and the stimulus presentation rate was set at 3.1/s for Study 1, and at a rate of 8.4/s for Study 2. An epoch time of 319.8 ms was used, yielding a sampling rate of 3202 Hz for Study 1 due to the recording system's fixed maximum 1024-point sampling size. In Study 2, the epoch time is 106.6 ms, resulting in a sampling rate of 9606 Hz. An onset recording delay of 3.4 ms was applied in order to maximize the capture of the sustained response.

For Study 1, each of the four sound level conditions was presented pseudo-randomly in two consecutive 1500-sweep blocks, for a total of 3000 sweeps per condition. Sweeps were presented in alternating polarity (for a total of 750 recorded sweeps at each polarity per block), and the eFFR was obtained by averaging the summed responses to the stimulus in each of the two polarities. The sFFR was similarly obtained by averaging the difference between the responses at each of the two polarities (Aiken & Picton, 2008).

Study 2 was similar, but two recording sessions were performed in order to obtain more data for subsequent machine learning analysis performed in Study 3, but also to enable an analysis of the test-retest stability of the responses. For each of the four vowels, each of the four sound level conditions was recorded in two consecutive 1500-sweep blocks, for a total of 3000 sweeps per condition per session. In total, 6000 sweeps were therefore collected for each unique condition per subject.

It should be noted that the effect of electromagnetic leakage and the possible contamination of recorded responses was previously investigated in our lab by replicating the recording conditions of Study 1 as described above, with the exception that the earphone was inserted

into an ear simulator (IEC 60318-4, G.R.A.S. RA0045) rather than the right ear of the subject, and the subject had earplugs inserted into both ears to prevent them from hearing any acoustic sounds (Laroche, 2010; Laroche et al., 2013). The spectrum of the eFFR showed no leakage at F0, and the sFFR showed no leakage about F1, thus indicating no meaningful electromagnetic leakage from the lab equipment to the electrodes (Laroche, 2010; Laroche et al., 2013).

4.2.4 Data handling and processing

All of the recorded amplitude data were exported to ASCII text files using the Bio-logic 'AEP to ASCII' (v. 1.2.1) software, and they were subsequently imported into MATLAB R2014a (v.8.3.0.532, by MathWorks, Natick, MA) for offline analysis. All FFRs were detrended by subtracting their respective means (Akhoun et al., 2008). The frequency domain data were then generated by taking the Discrete Fourier Transform (DFT) of the responses. Grand-average eFFR and sFFR waveforms were formed by summing the waveforms from all of the subjects and by dividing by the total number of waveforms summed. Grand-averages of the (non-RMS) amplitude spectra were similarly calculated by summing the amplitude spectra of all subjects and by dividing by the number of spectral series in the sum. Grand-average time and frequency domain plots of the eFFRs and sFFRs and the data requisite for their production were explicitly generated for each subject per condition, and grand-averages across all subjects were generated per condition as well. The amplitude of the spectral data for F0 and its harmonics were extracted from each FFR, and intra-subject grand-averages were saved for statistical testing purposes. The RMS-amplitudes of various harmonic combinations of interest were also calculated as the square root of the sum of the squared harmonic amplitudes for statistical testing purposes (e.g. in order to assess the growth of the harmonics associated with a formant).

The 300 ms /a/ vowel FFRs were all bandpass filtered offline from 65-1440 Hz (10th order Butterworth) prior to plot generation and subsequent statistical testing procedures. For the sake of the major peak analyses carried out in both studies, the eFFRs from Study 1were additionally bandpass filtered from 65-650 Hz and the eFFRs from Study 2 were lowpass filtered with a cutoff at 650 Hz (10th order Butterworth) in order to remove high frequency noise from the waveforms and to minimize the number of corrupted or broken peaks encountered as a result. All filtering was performed using Matlab's 'filtfilt' function to prevent introducing filtering delays or changing the phase of components.

4.2.5 Data analysis

All statistical analyses were conducted using SPSS (v.24, IBM, Armonk, NY). An alpha value of 0.05 was used throughout. In Study 1, repeated-measures analysis of variance tests (RMANOVAs) were conducted to test for intra-subjects sound level differences in selected harmonic spectra and their combined RMS-amplitudes, with main effects of level being reported along with η_p^2 to provide a measure of effect size (Richardson, 2011). Wherever significant effects were observed based on Wilks' lambda test statistic, pairwise comparison tests were performed between each sound level-pairing, and the p-values were adjusted using the Holm-Bonferonni correction (Holm, 1979; Park, Cho, & Ki, 2009). Details of this procedure, which maintains the selected type I error rate, α, while increasing power (i.e. decreasing the type II error rate) relative to the original Bonferroni procedure, can be found in Appendix E. Here, only one-sided t-tests were conducted, as it was hypothesized that all harmonic components and the selected combinations would increase with increasing level. The features were selected based on the literature and piloting, and included F0, F0&H2 combined, F0-H6 (indicating all six harmonics in combination), H2, H2-H6, and H3-H6 among the eFFRs. The sFFR feature set was comprised of F0, H2, H6-H8 (a global F1 measure), H7 (F1), H11-H13 (a global F2 measure), and H12 (F2). A measure of the overall sFFR growth – F0-H12 – was also included.

The statistical analyses of the harmonic spectra in Study 2 were virtually identical to that of Study 1 as outlined above, with RMANOVAs being conducted to test for intra-subject sound level differences; however, a mixed design was used here with 'sex' acting as the between-subjects factor. Wherever sex-level interaction effects were observed, simple effects of level and of sex are subsequently reported, with pairwise comparisons between sound level pairs being performed where simple effects of level are observed, and pairwise comparisons between sexes being performed for each sound level where simple effects of sex are observed. Whenever sex-level interaction effects were not observed, tests on the main effects of level and the main effects of sex were conducted, with subsequent pairwise comparisons being performed where appropriate.

The spectral features analysed in Study 2, which were selected based on the results of Study 1 and the review of the literature, consisted of F0, H2, and H2-H6 for the eFFRs, and the centre F1 and F2 harmonics for each vowel, as well as the two global measures of formant growth, which included one harmonic on either side of the centre formant frequencies (e.g. H6-H8 for the /a/ vowel with F1 centred about H7) for the sFFRs. The pairwise comparisons between level conditions in Study 2 were all one sided t-tests aside from the tests performed on F0 as extracted from the eFFR amplitude spectra, as it was hypothesized that

all other harmonics and the selected combinations thereof would exhibit increasing growth with increases in sound level. The preliminary results from Study 1 clearly indicated that F0 did not exhibit the expected monotonic growth, and so pairwise comparisons were two-sided in Study 2 to reflect the hypothesis that differences would be observed, but that they might be either increasing or decreasing with increasing level.

Analysis in the time-domain was restricted to aspects of the steady-state response, and no treatment of the transient (e.g. onset) responses was undertaken. This is due to the fact that such responses do not appear to be reliable in individual waveforms in these data, and that such responses are also unlikely to be reliable in aided conditions due to processing-related delays. Here, major peaks were first selected from each subject's grand-average eFFR by a peak-picking script written in MATLAB and were subsequently verified visually. The algorithm functioned by selecting local maxima in Study 1, since these peaks were deemed to be the most robust in the grand-average waveforms, and by picking local minima in Study 2 for the same reason. Peaks that were selected for but that were not within ±2 ms of the grand-average mean for those peaks were discarded in order to avoid the frequently occurring minor peaks that typically occurred 5 ms apart from the major peaks (which was verified via filtering to be primarily due to the 200 Hz component). Where two or more local optima that occurred within fractions of a millisecond of one another were selected by the algorithm, a visual inspection of the peak was performed in order to select the peak. If two data points were equally likely candidates and the similar local optima appeared to be caused by a very minor deflection (i.e. a broken peak), then the morphology was attributed to noise and the average of the two points was taken. In the rare case (<1% of the time) where a peak was unable to be identified, interpolation was performed by taking the midpoint between the two adjacent peaks. Both peak latencies and amplitudes were recorded, although amplitudes were only used to establish norms and to provide guidance for Study 3, and no statistical testing was undertaken in this regard here.

In order to analyze the peak latencies with respect to systematic shifts between level conditions, a normalization process similar to that described by K. L. Johnson et al. (2008) in their analysis of the FFR peaks of voiced consonant-vowel stop syllables was performed. The normalization was first performed by computing a grand mean (GM) latency for each major peak across all four sound level conditions, and then by subtracting said mean from each subject's respective peak latency for each sound level condition (i.e. by taking $\mathrm{Lat}_{\mathrm{SUB}}$ – $\mathrm{Lat}_{\mathrm{GM}}$ for all sound levels and all peaks). This results in earlier peaks taking on negative values, later peaks taking on positive values, and those closest to the mean being nearest to zero. Each subject's normalized latency per condition provides a simple numerical indication of how their peak latencies change with changing level, enabling a clear and common inter-

pretation of the latency regardless of whether it was for a peak occurring tens of milliseconds or few hundred milliseconds after stimulus onset.

RMANOVAs were performed on the peaks of the eFFRs of the subjects' normalized latencies to test for systematic differences across peaks as a function of sound level in Study 1. In Study 2, mixed-design RMANOVAs using 'sex' as a between-subjects factor were performed in order to test for sex-level interactions; however no such interactions were observed. Wherever significant main effects of level were observed, follow up RMANOVAs on the absolute latencies for each of the major peaks were performed. Where effects of level were observed on a peak's latencies, pairwise comparisons were subsequently performed via one-sided t-tests since it was hypothesized that latency would decrease with increasing level. The Holm-Bonferroni correction was again employed to account for multiple testing. Wherever significant main effects of sex were observed, they were reported, along with the mean latency difference between the sexes. No additional follow-up tests of sex effects on the individual major peaks were performed since overall mean latency differences were of greatest interest beyond the assessment of interaction effects here.

Response latencies were calculated for all stimuli at all four sound levels via the cross-correlation of the respective grand-average eFFRs with an inverted version of the corresponding stimulus envelope, obtained using the Hilbert transform that was lowpass filtered (Fc=2000 Hz) and down-sampled appropriately to match the sampling rate of each eFFR (Akhoun et al., 2008; Laroche, 2010). The latency value that corresponded to the maximum correlation value in the known FFR conduction delay range (~5-15 ms) was selected. Corrections accounting for the 1 ms delay in acoustic transmission up to the insert-earphone were applied throughout. Response latencies from Study 2 were additionally corrected to account for the 3.4 ms delay in the onset of the recording relative to the onset of the stimulus.

Correlations between subaverages were also calculated in both studies. The correlations were performed on the bandpass filtered signals (10th order Butterworth from 65-1440 Hz) in Study 1, and on the unfiltered signals in Study 2. For each stimulus and for each correlative measure of interest, the correlations were calculated on the eFFRs and sFFRs in both the time domain and the frequency domain (via the amplitude spectra), and ranges and standard deviations are provided throughout where appropriate. Mean within-level intra-subject correlations were calculated between the two subaverages of each subject at each level condition in Study 1 and between the 'test' and 'retest' grand-averages of each subject at each level condition in Study 2. This measure provides insight into the level-dependency of the FFR's stability in the time and frequency domains. Mean intra-subject correlations were also computed by averaging across level conditions in order to provide a general sense of the intra-subject stability of the FFR across sound levels.

In Study 1, mean inter-subject correlations were also computed using each subject's grand-average FFRs per level condition. The same was performed with each vowel stimulus in Study 2, but rather than tabulating the inter-subject correlations outright (similar to those in Study 1) the percentage of the occasions that an individual's 'test' subaverage yielded the greatest correlation with their own 'retest' subaverage rather than with another subject's was recorded.

Finally, mean peak-to-peak interval times were calculated between the previously identified major peaks of each subject in both studies, and the average interval times were computed for each level condition of each stimulus. The standard deviation amongst the subjects was also computed here. This was performed to provide a sense of the temporal precision of the encoding of the pitch period (i.e. 1/F0) for each stimulus and level.

4.3 Results

The results of Study 1 are first presented below, beginning with the frequency domain analyses, and then followed by the time-domain analyses. The results of Study 2 are then presented using the same format, with the added summarization of the test-retest analysis. Note that in all plots corresponding to the grand-average FFRs, the bottom subplots representative of the stimulus or its envelope (as indicated) have been normalized to 1 throughout for visualization purposes.

4.3.1 Study 1

Frequency domain analysis

The grand-average eFFR in response to the 300 ms /a/ vowel is shown in Fig. 4.2, with the left column displaying the time-domain waveform and the right column displaying the frequency-domain amplitude spectra. The bottom row subplots are of the inverted envelope of the stimulus itself in each of the respective domains. Although some morphological differences between level conditions are clearly visible in the time domain, the changes in the amplitude spectra are less evident.

Figure 4.3 shows the mean amplitude spectra for the first six harmonics of F0 (F0 to H6) and provides a clearer indication of their respective level-related growth functions. A tabulation of the tests performed on the selected harmonics and their combinations is provided in Table 4.2. Although significant effects of level in F0 amplitude were observed ($F(3,9)=7.770$, $p=.007$), none of the pairwise comparisons between levels attained significance here. This was surprising since monotonic increasing growth was hypothesized; however, Fig. 4.3 clearly

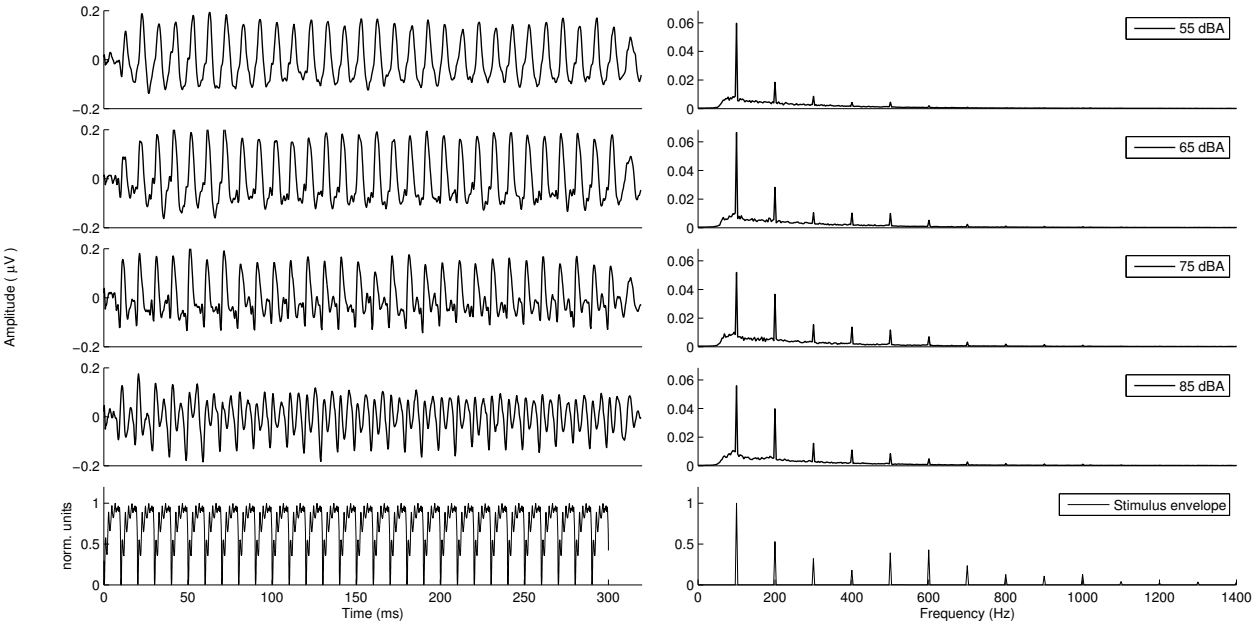

Figure 4.2: Grand-average eFFR to a 300 ms /a/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA in both the time domain (left), and frequency domain via the averaging of the individual amplitude spectra (right). The bottom panel shows the normalized envelope characteristics of the stimulus as calculated using the Hilbert transform. [Study 1].

Table 4.2: Statistical analysis of the eFFR to a 300ms /a/ vowel at selected harmonics and their combinations. Bold numbers indicate statistical significance. [Study 1].

	Main effect of level		**Pairwise comparisons**					
Feat.	**F-stat.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
F0	**F(3,9)=7.770, p=.007**	**.721**	.0553	.8875	.5685	.999	.678	.427
F0&H2	**F(3,9)=6.037, p=.015**	**.832**	.010	.281	.3335	.9685	.5175	.3895
F0-H6	**F(3,9)=6.167, p=.015**	**.673**	**.0055**	.1155	.279	.9135	.488	.399
H2	**F(3,9)=14.813, p=.001**	**.832**	**.0005**	**.0005**	**.002**	**.004**	.0505	.2425
H2-H6	**F(3,9)=23.974, p=.0005**	**.889**	**.0005**	**.0005**	**.001**	**.0005**	**.0485**	.396
H3-H6	**F(3,9)=39.751, p=.0005**	**.930**	**.0005**	**.0005**	**.0005**	**.0005**	.0795	.898

indicates a non-monotonic function of level for F0. Similarly to F0, significant effects of level were observed for the combined measures of F0&H2 and F0-H6 (F(3,9)=6.037, p=.015 and F(3,9)=6.167, p=.015, respectively) while none of the pairwise comparisons attained significance themselves aside from the 55 vs. 65 dBA F0-H6 test. Significant effects of level were observed for H2 (F(3,9)=14.813, p=.001), and all pairwise comparisons attained significance aside from the 65 vs. 85 and 75 vs. 85 dBA tests. Significant effects of level were also observed for the H2-H6 and H3-H6 measures (F(3,9)=23.974, p=.0005 and F(3,9)=39.751, p=.0005, respectively). All pairwise comparisons were significant for the H2-H6 measure other than the 75 vs. 85 dBA test. The H3-H6 measure failed to attain significance on both the 65 vs. 85 and 75 vs. 85 dBA pairwise comparisons, but attained significance on all others.

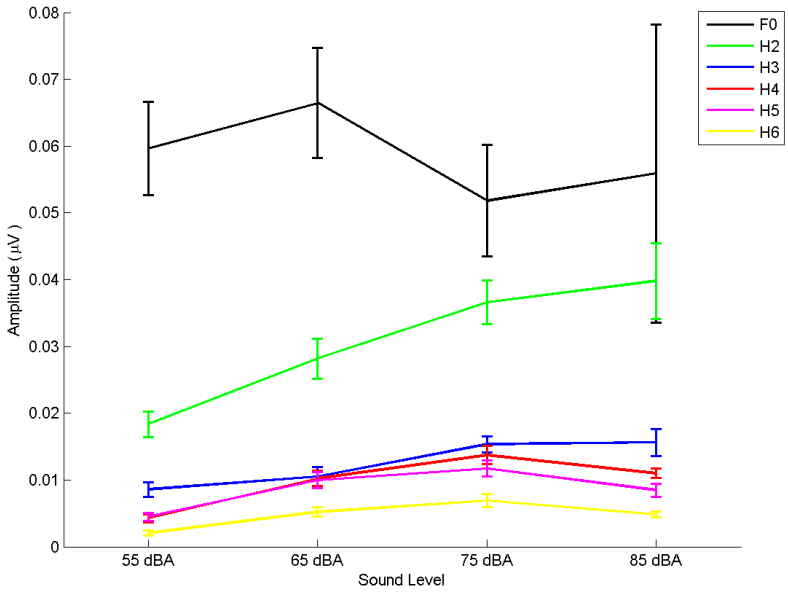

Figure 4.3: The mean eFFR amplitude spectra (in microvolts) of harmonics F0 to H6 for the 300 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 1].

The grand-average sFFRs are shown in Fig. 4.4. The layout here is identical to that of the eFFR in Fig. 4.2 aside from the fact that the bottom panel is now representative of the stimulus proper and not its envelope. Due to the higher frequency components and the smaller magnitude of the measured response (see Section 2.2.2), morphological differences in the sFFR are not as visually obvious in the time-domain panel as they are for the eFFR. Evident, however, are the growth trends among the harmonics about the first formant (H5-H7) with increasing level. These growth trends, along with those of all of the other individual sFFR harmonics of interest can be more clearly observed in Figs. 4.5 to 4.7, which show the low (F0 to H4), mid (H5 to H8) and high (H9 to H13) harmonics, respectively.

Results of the statistical tests carried out on the selected harmonics and harmonic combinations are tabulated in Table 4.3. Differences in F0 amplitude were not observed ($F(3,9)=2.363$, $p=.139$), and it is not clear that the measures of F0 at 55 and 65 dBA are not merely representative of noise (see Fig. 4.4). However, all of the other main effects of level tests attained significance. As hypothesized, the measures associated with the growth of F1 (namely H7 and H6-H8) attained significance ($F(3,9)=10.059$, $p=.003$ and $F(3,9)=15.736$, $p=.001$, respectively) and all of the pairwise comparisons were statistically significant as well. Even though the high-cut filter was set at 1000 Hz, which is below the second formant region, the grand-average sFFR revealed visually obvious growth and the statistical tests of H12 and H11-H13 confirmed this ($F(3,9)=5.941$, $p=.016$, $F(3,9)=4.685$, $p=.031$, respectively). Both of these measures attained significance for all pairwise comparisons other than on the 55

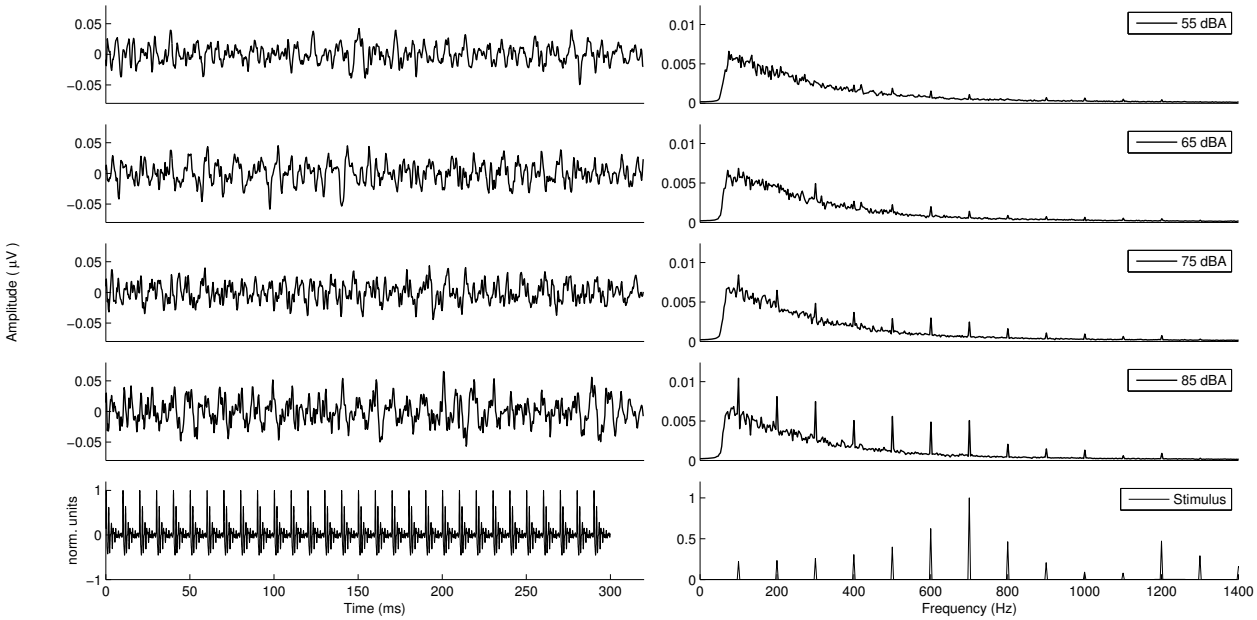

Figure 4.4: Grand-average sFFR to a 300 ms /a/ vowel with F0=100 Hz, F1 centred at 700 Hz, and F2 centred at 1200 Hz. The stimulus was presented at 55, 65, 75, and 85 dBA , and the response is shown here in both the time domain (left), and frequency domain via the averaging of the individual amplitude spectra (right). The bottom panel shows the normalized characteristics of the stimulus. [Study 1].

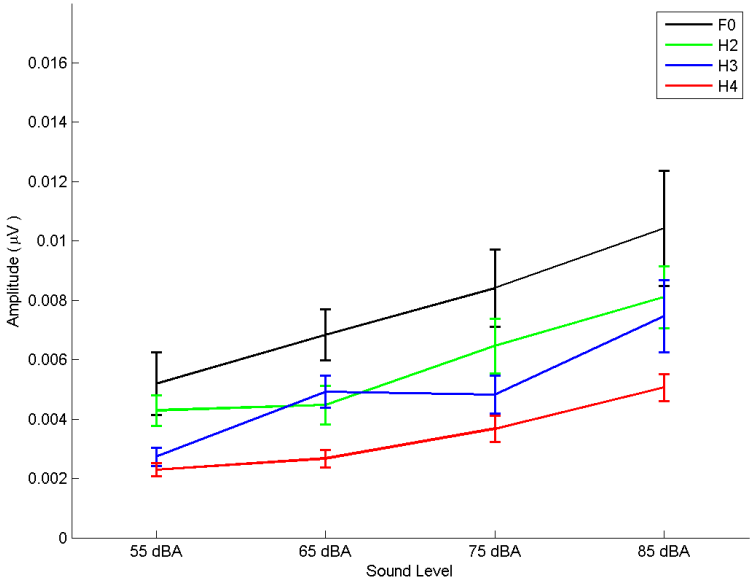

Figure 4.5: The mean sFFR amplitude spectra (in microvolts) of the low harmonics (F0 to H4) for the 300 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 1].

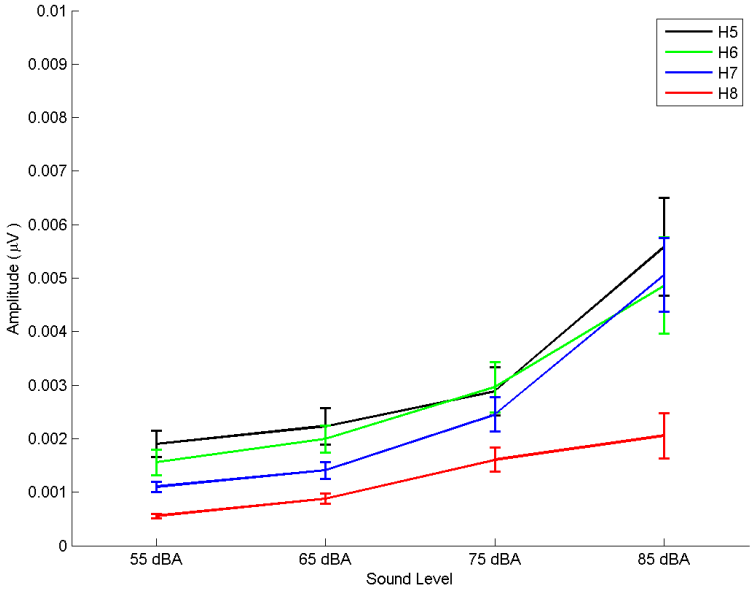

Figure 4.6: The mean sFFR amplitude spectra (in microvolts) of the mid harmonics (H5 to H8) for the 300 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 1].

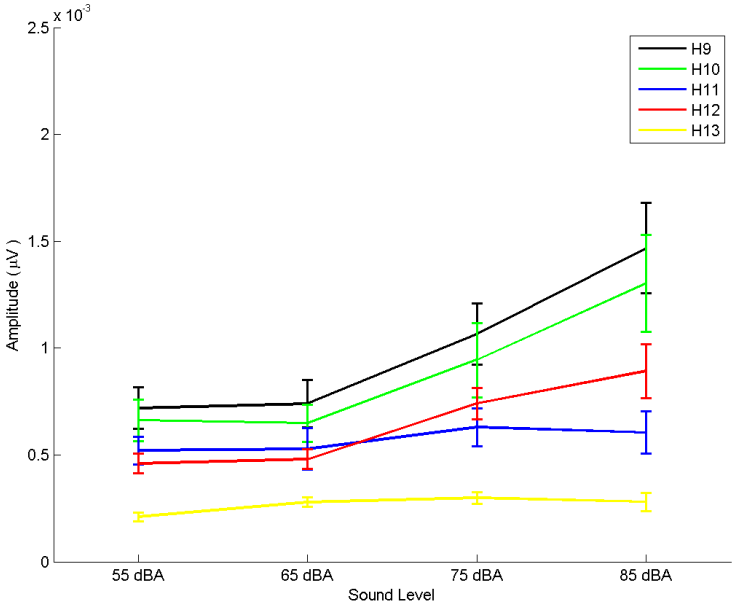

Figure 4.7: The mean sFFR amplitude spectra (in microvolts) of the high harmonics (H9 to H12) for the 300 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars represent the SEM. [Study 1].

Table 4.3: Statistical analysis of the sFFR to a 300ms /a/ vowel at selected harmonics and their combinations. Bold numbers indicate statistical significance. [Study 1].

	Main effect of level		**Pairwise comparisons**					
Feat.	**F-stat.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
F0	F(3,9)=2.363, p=.139	.441	NA	NA	NA	NA	NA	NA
H2	**F(3,9)=7.864, p=.007**	**.724**	.375	**.0015**	**.001**	**.0065**	**.008**	.078
H6-H8	**F(3,9)=15.736, p=.001**	**.840**	**.007**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
H7	**F(3,9)=10.059, p=.003**	**.770**	**.039**	**.0015**	**.0005**	**.003**	**.0005**	**.0005**
H12	**F(3,9)=5.941, p=.016**	**.664**	.3205	**.0015**	**.008**	**.0005**	**.0055**	.1085
H11-H13	**F(3,9)=4.685, p=.031**	**.610**	.186	**.001**	**.003**	**.0095**	**.004**	.0695
F0-H12	**F(3,9)=8.945,p=.005**	**.749**	**.0095**	**.0005**	**.0005**	**.001**	**.0005**	**.0015**

Table 4.4: The mean (SD) interpeak intervals of the eFFR to a 300 ms /a/ vowel for each level condition. [Study 1].

Sound level (dBA)	**Mean interpeak interval (ms)**
55	10.0 (.5)
65	10.0 (.3)
75	10.0 (.3)
85	10.0 (.3)

vs. 65 and 75 vs. 85 dBA tests. A combined measure of all of the harmonics from F0-H12 attained significance (F(3,9)=8.945, p=.005), and did so among all pairwise comparisons as well.

Time domain analysis

The grand-averages of the time-domain representations for both the eFFR and sFFR are shown in Figs. 4.2 and 4.4. Figure 4.8 displays the relative normalized latency of the major peaks at each sound level. Notice that the lower the sound level, the greater the normalized latency tends to be, and vice-versa. Table 4.4 indicates the mean interpeak interval within subjects using the same major peak data and reveals a very precise tracking of the pitch period of 10 ms (1/F0) across level.

The visually apparent trends in Fig. 4.8 are as expected based on the AEP literature (see Section 2.3) and it is not surprising that the RMANOVA on the normalized means at each sound level revealed a significant effect of level (F(3,7)=186.411, p<.0005, η_p^2 = .988). Additional tests were carried out on each of the 30 major peaks. The results of these tests, shown in Table 4.5, are all statistically significant – both for the RMANOVAs and for the consequent pairwise comparisons.

The amplitudes of these same major peaks are displayed in Fig. 4.9. Interestingly, the 55 and 85 dBA peaks are similar in amplitude profile (with the 85 dBA clearly possessing the lowest mean amplitude), and the 65 and 75 dBA peaks are very similar as well. Recall that

Table 4.5: Results of RMANOVAs on the latencies of each major peak of the eFFR to a 300 ms /a/ vowel as a function of sound level, and the subsequent pairwise comparisons performed between levels. Note that all tests obtain statistical significance (indicated in bold). [Study 1].

	Main effect of level		**Pairwise comparisons between sound levels (p-val.)**					
Peak #	**F(3,8)**	**p-val.**	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
1	**105.856**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
2	**89.703**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
3	**68.363**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
4	**81.211**	**.0005**	**.0005**	**.0005**	**.0005**	**.002**	**.0005**	**.0005**
5	**165.902**	**.0005**	**.001**	**.0005**	**.0005**	**.0075**	**.0005**	**.0005**
6	**138.256**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
7	**118.103**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
8	**158.147**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
9	**52.974**	**.0005**	**.0245**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
10	**206.840**	**.0005**	**.0005**	**.0005**	**.0005**	**.0015**	**.043**	**.0005**
11	**42.218**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
12	**73.335**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
13	**278.168**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
14	**66.170**	**.0005**	**.013**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
15	**188.494**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
16	**91.600**	**.0005**	**.0017**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
17	**96.501**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
18	**21.000**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
19	**89.547**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
20	**61.616**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
21	**168.248**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
22	**142.060**	**.0005**	**.002**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
23	**118.733**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
24	**251.875**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
25	**122.211**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
26	**81.929**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
27	**182.187**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
28	**123.677**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
29	**106.701**	**.0005**	**.0005**	**.0005**	**.0005**	**.002**	**.0005**	**.0005**
30	**44.820**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.001**

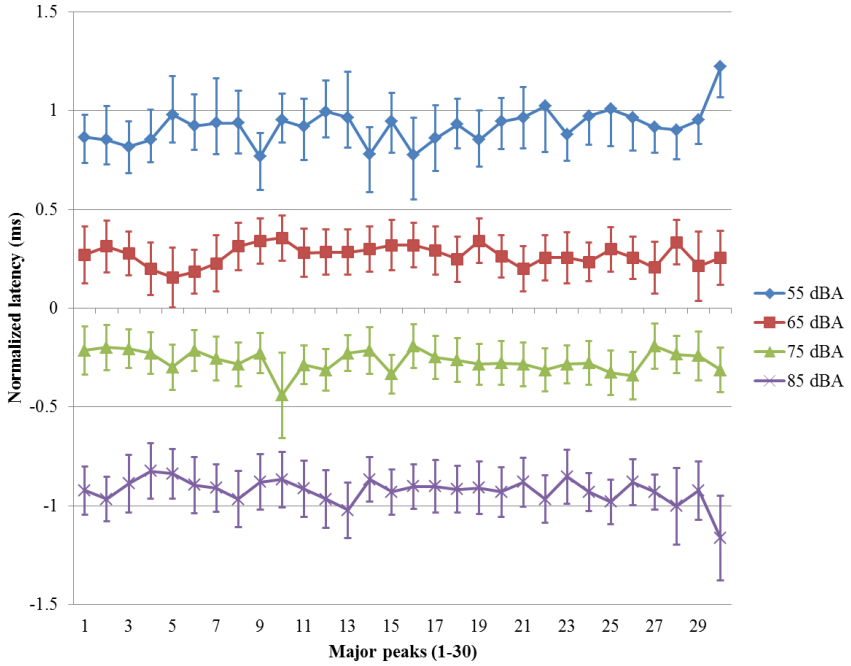

Figure 4.8: Relative (normalized) latency (ms) of the major peaks present in the eFFRs to a 300 ms /a/ vowel as a function of sound level. The more positive the latency value, the later the peak occurred and vice-versa. The error bars represent the SEM. [Study 1].

no statistical tests were performed here since the major peak amplitude trends were largely of interest in order to determine if they should be utilized as input features in the machine learning models examined in Study 3.

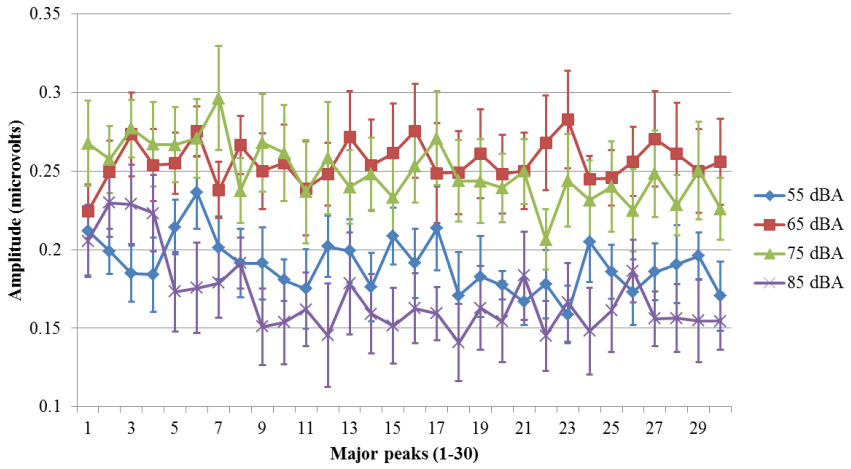

Figure 4.9: The mean major peak amplitudes (microvolts) of the eFFR to the 300 ms /a/ vowel are plotted as a function of level, with the error bars indicating the SEM. [Study 1].

Finally, the response latencies per sound level are indicated in Table 4.6. As hypothesized, they decrease with increasing level, and they lie in the range expected based on the ABR and FFR literature (see Section 2.3 and Appendix B).

Table 4.6: Response latencies (eFFR) for the 300 ms /a/ vowel. [Study 1].

Sound level (dBA)	Latency (ms)
55	8.4
65	8.7
75	9.3
85	10.2

Correlation results

The results of the correlation analysis on the eFFR and sFFR subaverages in both the time and frequency domains are summarized in Tables 4.7 to 4.9. These results indicate that strong correlations exist between the eFFR subaverages in both the time domain and frequency domain as well as among the frequency domain representations of the sFFRs. The sFFR waveforms demonstrate weak correlations (if any whatsoever). The mean within-level inter-subject correlations appear to be slightly weaker than those of the intra-subject instance among the eFFR waveforms, especially with increasing level, which appears to hold in the frequency domain as well. However, the respective sFFR frequency-domain correlations appear to be stronger between subjects than within subjects.

Table 4.7: Mean (SD) within-level intra-subject Pearson correlations over subaverages for the 300 ms /a/ vowel. [Study 1].

Sound level (dBA)	eFFR time domain	sFFR time domain	eFFR frequency domain	sFFR frequency domain
55	.64 (.17)	.03 (.03)	.89 (.06)	.69 (.05)
65	.67 (.16)	.06 (.06)	.90 (.04)	.70 (.03)
75	.58 (.21)	.02 (.02)	.88 (.06)	.70 (.03)
85	.61 (.23)	.14 (.14)	.88 (.07)	.73 (.04)

Table 4.8: Mean (SD) intra-subject Pearson correlations over subaverages in response to a 300 ms /a/ vowel, here averaged across sound levels. [Study 1].

eFFR (time)	sFFR (time)	eFFR (freq.)	sFFR (freq.)
.63 (.04)	.06 (.09)	.89 (.04)	.71 (.03)

Table 4.9: Mean (SD) within-level inter-subject Pearson correlations for the 300 ms /a/ vowel. [Study 1].

Sound level (dBA)	eFFR (time)	sFFR (time)	eFFR (freq.)	sFFR (freq.)
55	.59 (.13)	.00 (.08)	.90 (.04)	.82 (.02)
65	.57 (.14)	.01 (.07)	.89 (.05)	.80 (.03)
75	.37 (.23)	.00 (.09)	.85 (.07)	.80 (.02)
85	.26 (.26)	.04 (.10)	.82 (.09)	.80 (.03)

4.3.2 Study 2

A preliminary examination of some of these results was reported on in Heffernan, Dajani, and Giguère (2017).

Frequency domain analysis

The grand-average eFFRs and sFFRs for each of the 100 ms /a/, /ɔ/, /U/, and /u/ vowels (displayed in Figs. 4.10, 4.16, 4.22 and 4.28 and Figs. 4.12, 4.18, 4.24 and 4.30, respectively) are identical in layout to those of Study 1 aside from the fact that each plot shows grand-averages of both male (black) and female (red) subjects. Below each respective figure, plots of the relevant harmonic spectra and tabulations of the statistical tests of the selected features are shown. The eFFR tests, which were informed by the results of Study 1, include tests for statistical differences in the spectral amplitudes of F0, the second harmonic (H2), and the second through to sixth harmonics combined (H2-H6). The sFFR tests, which were also informed by the results of Study 1, include tests of the spectral amplitude of F1 and F2 for each stimulus - both the centre harmonic of the formant, and the combination of itself and the two harmonics on either side of it. From Table 4.1, the centre harmonics and three-harmonic combinations for the F1 measures are respectively H7 and H6-H8 for vowel /a/, H6 and H5-H8 for /ɔ/, H5 and H4-H6 for the /U/ vowel, and H3 and H2-H4 for vowel /u/. Similarly, the F2 measures are respectively H12 and H11-H13 for the /a/ and /U/ vowels, H9 and H8-H10 for the /ɔ/ and /u/ vowels.

The 100 ms /a/ vowel's grand-average eFFR (Fig. 4.10) appears to reveal very similar responses between sexes in both the time and frequency domains across level. The minor yet evident differences seem to indicate that the amplitude spectra of certain harmonics is slightly larger among females, and that the waveforms of the females tend to lead those of the males, as can be seen by attending to the major negative peaks that are spaced apart by roughly 10 ms. Despite these seeming differences, no sex-level interaction effects were observed in the eFFR frequency domain tests, as can be seen in Table 4.10. Consequently, main effects of level were tested and all three measures of F0, H2, and H2-H6 attained significance ($F(3,18)=18.929$, $p=.0005$; $F(3,18)=27.430$, $p=.0005$; and $F(3,18)=50.259$, $p=.0005$, respectively). The pairwise comparisons on F0 means attained significance on all but the 55 vs. 75 dBA tests, and the pairwise comparisons on H2 means all attained significance. Both of these results can be corroborated by visual inspection of Fig. 4.11, which shows the growth functions of the individual eFFR harmonics. All pairwise comparisons attained significance on the H2-H6 measure of 'harmonic richness'.

Main effects of sex were also tested (see Table 4.10). F0 did not exhibit significant

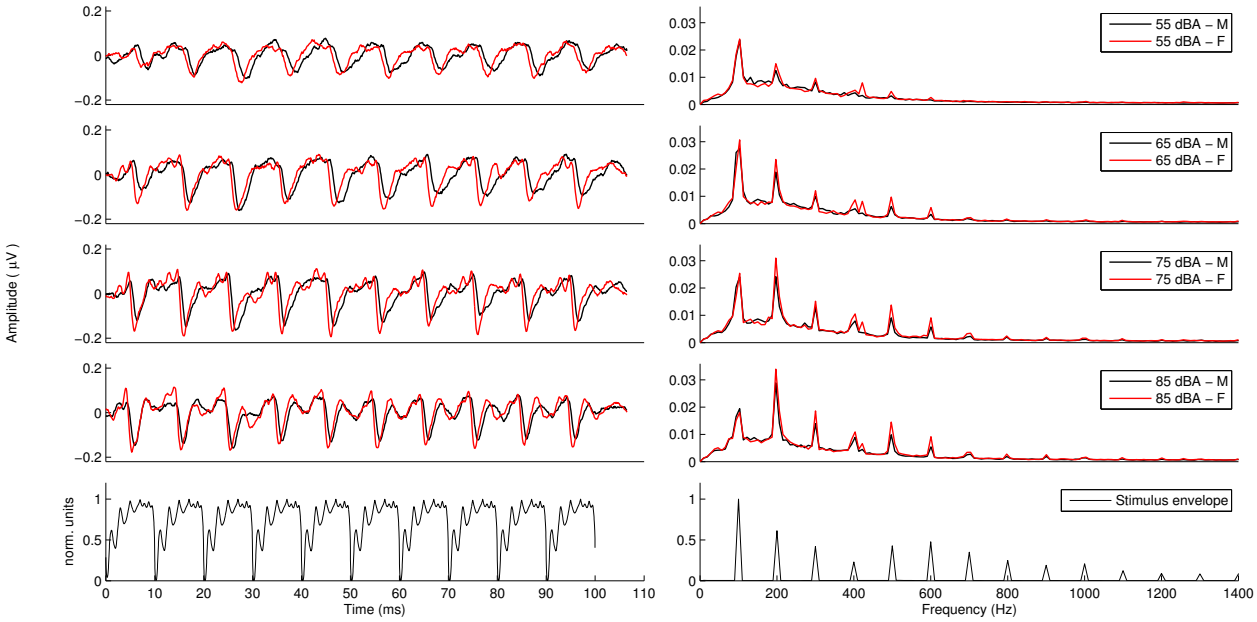

Figure 4.10: Grand-average eFFR of males (black) and females (red) to a 100 ms /a/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the normalized envelope characteristics of the stimulus as calculated using the Hilbert transform. [Study 2].

Table 4.10: Statistical analysis of the eFFR to a 100ms /a/ vowel for F0, H2, and H2-H6 features. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-vals)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
F0	F=.968, p=.429	**F=18.929, p=.0005**	**.759**	**.003**	.650	**.016**	**.020**	**.0005**	**.0005**	F=.141, p=.711	.007
H2	F=.379, p=.770	**F=27.430, p=.0005**	**.821**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.004**	**F=5.270, p=.033**	**.209**
H2-H6	F=.927, p=.44	**F=50.259, p=.0005**	**.893**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.002**	**F=11.235, p=.003**	**.360**

effects of sex here (F(1,20)=.141, p=.711); however, both the H2 and H2-H6 measures did (F(1,20)=5.270, p=.033; and F(1,20)=11.235, p=.003, respectively).

The grand-average sFFRs for the 100 ms /a/ vowel shown in Fig. 4.12 reveal the expected growth of the harmonics in the vicinity of F1 (700 Hz), and seem to show some growth about F2 (1200 Hz) as well, although the spectral amplitudes are relatively small in the higher frequency region, as expected due to the low-pass like nature of the FFR that results from the consistent drop-off in the upper limits of phase-locking as a signal ascends from the periphery to the cortex. Figs. 4.13 to 4.15 provide a clearer picture of the level-related growth functions for each of the harmonics in the low (F0 to H4), mid (H5 to H8), and high (H9 to H13) harmonic ranges.

The aforementioned visual trends are corroborated by the results of the statistical tests compiled in Tables 4.11 to 4.13. The results are extremely similar to those of Study 1 and

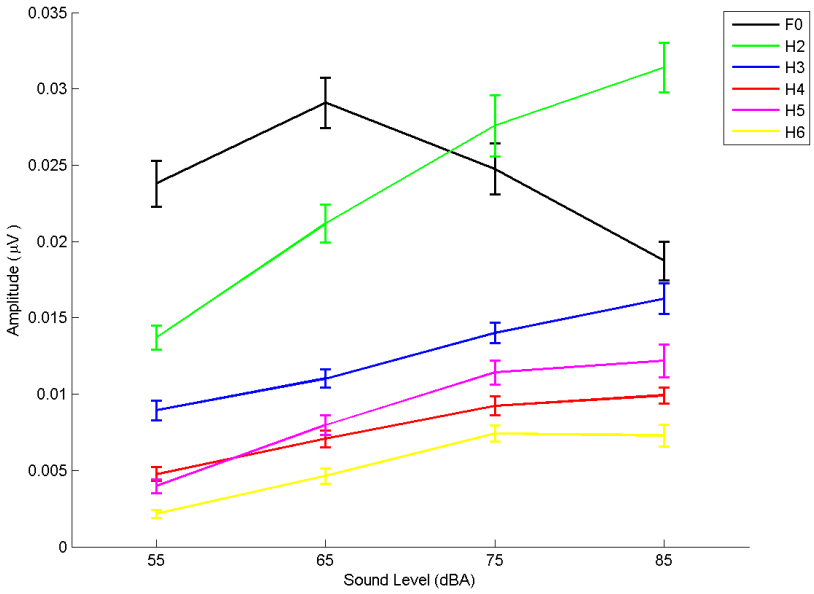

Figure 4.11: The mean eFFR amplitude spectra (in microvolts) of harmonics F0 to H6 for the 100 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

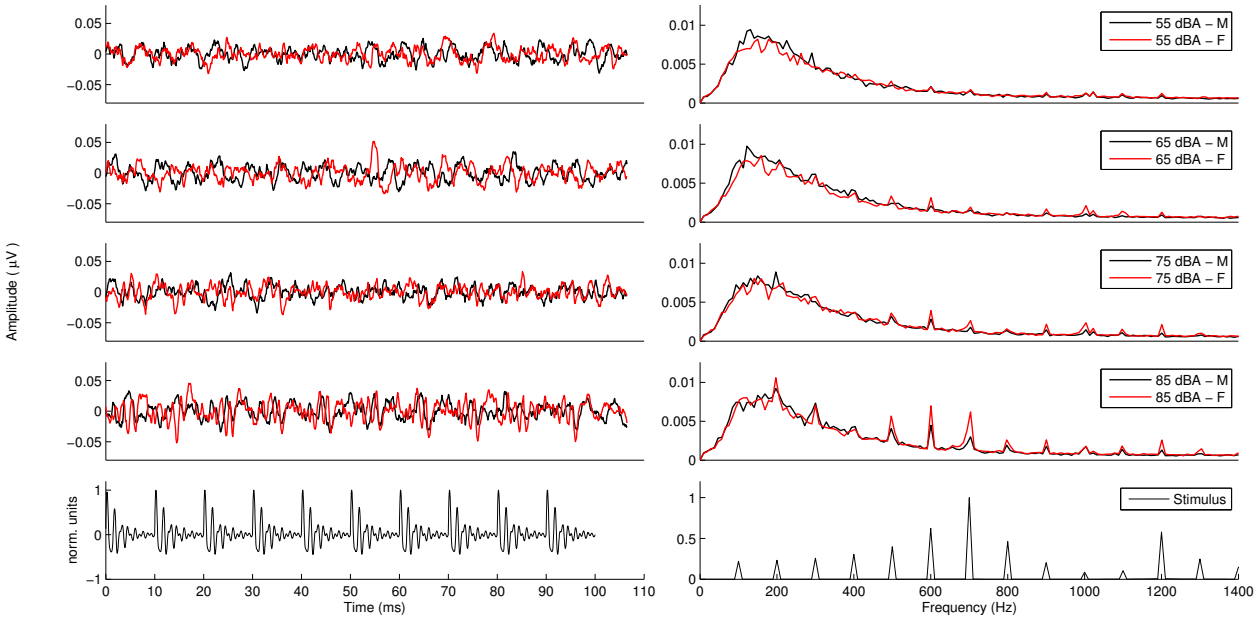

Figure 4.12: Grand-average sFFR of males (black) and females (red) to a 100 ms /a/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the characteristics of the stimulus, with F1 centred at 700 Hz, and F2 centred at 1200 Hz. [Study 2].

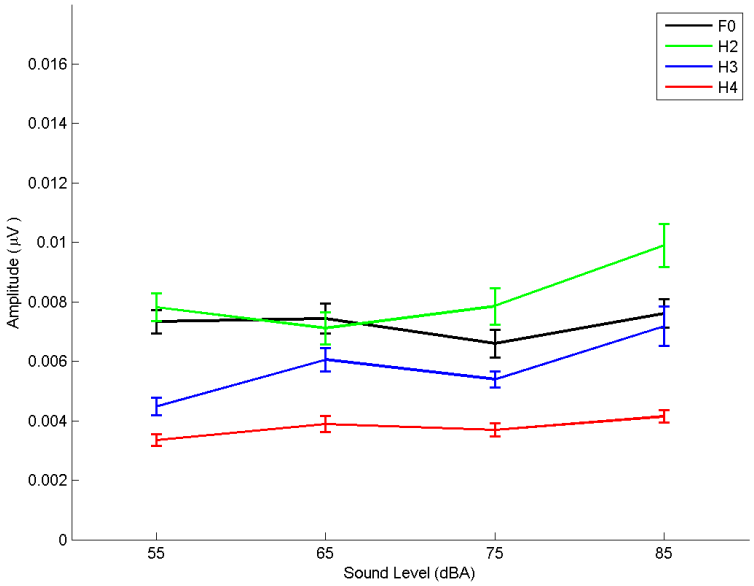

Figure 4.13: The mean sFFR amplitude spectra (in microvolts) of the low harmonics (F0 to H4) for the 100 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

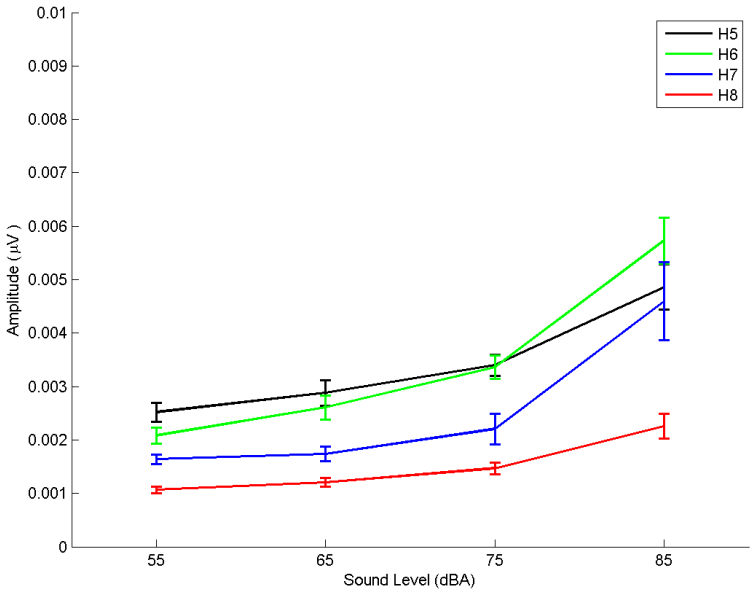

Figure 4.14: The mean sFFR amplitude spectra (in microvolts) of the mid harmonics (H5 to H8) for the 100 ms /a/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

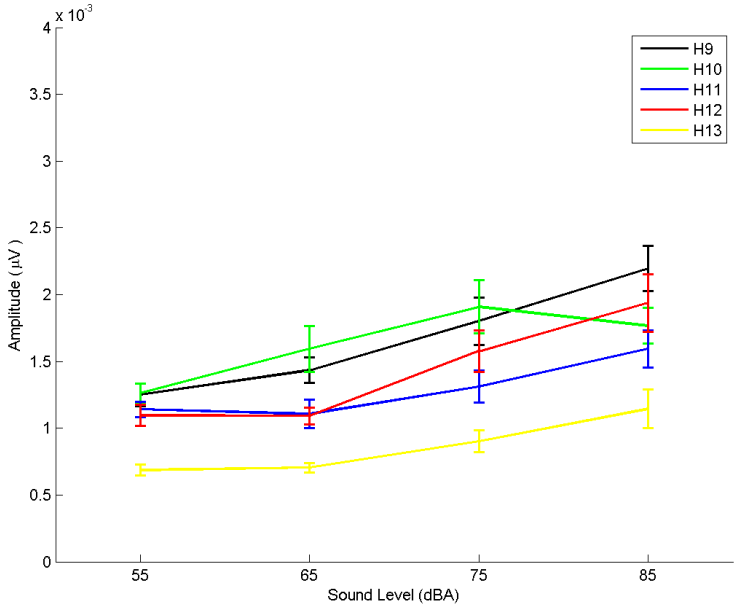

Figure 4.15: The mean sFFR amplitude spectra (in microvolts) of the high harmonics (H9 to H12) for the 100 ms /a/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

we again see that the more global measure of F1 growth, namely H6-H8, yields significance; however in this instance a sex-level interaction was detected (F(3,18)=4.554, p=.015) as indicated in Table 4.11). As a result of this significant interaction effect, the simple effects of level were tested for each sex with both the male and female groups achieving significance here (F(3,18)=6.869, p=.003 and F(3,18)=27.647, p=.0005, respectively). All pairwise comparisons were significant aside from the 55 vs. 65 dBA test within males. The simple effects of sex were also computed here (see Table 4.12), yielding significance for the 65 (F(1,20)=5.644, p=.028), 75 (F(1,20)=9.925, p=.005), and 85 dBA (F(1,20)=12.237, p=.002) levels, but not for the 55 dBA level (F(1,20)=.021, p=.886).

Table 4.11: Statistical analysis of the sFFR to a 100ms /a/ vowel (see Tables 4.12 and 4.13 as well). The F1 (H7) and F2 (H12) centre harmonics were tested for increasing growth with increasing level, as well as their more global formant measures of H6-H8, and H11-H13, respectively. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Simple effect of level**		**Pairwise comparisons between sound levels (p-values)**					
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
H6-H8 (F)	**F=4.554 p=.015**	**F=27.647, p=.0005**	**.822**	**.001**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
H6-H8 (M)	**F=4.554 p=.015**	**F=6.869, p=.003**	**.534**	.5005	**.0095**	**.001**	**.0025**	**.0005**	**.002**
H12 (F)	**F=5.743 p=.006**	**F=15.542, p=.0005**	**.721**	.369	**.0005**	**.0005**	**.0005**	**.0005**	**.0085**
H12 (M)	**F=5.743 p=.006**	F=.866, p=.477	.126	NA	NA	NA	NA	NA	NA

Table 4.12: Simple effects of sex within each level condition for the sFFR to a 100 ms /a/ vowel (corresponding to the sex-level interactions noted in Table 4.11). [Study 2].

	55 dBA		**65 dBA**		**75 dBA**		**85 dBA**	
Feat.	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2
H6-H8	F=.021, p=.886	.001	**F=5.644, p=.028**	**.220**	**F=9.925, p=.005**	**.332**	**F=12.237, p=.002**	**.380**
H12	F=3.940, p=.061	.165	**F=10.330, p=.004**	**.341**	**F=27.031, p=.0005**	**.575**	**F=14.598, p=.001**	**.422**

Table 4.13: Main effects results for the non-significant sex-level interactions of the sFFR to a 100 ms /a/ vowel. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
H7	F=2.129, p=.132	**F=9.721, p=.0005**	.618	.281	.027	**.0005**	.0305	**.0005**	**.0005**	**F=5.426, p=.030**	**.213**
H11-H13	F=2.134, p=.131	**F=13.471, p=.0005**	**.692**	**.485**	**.001**	**.0005**	**.0005**	**.0005**	**.0005**	**F=16.430, p=.001**	**.451**

Another sex-level interaction effect was observed for H12 (F(3,18)=5.74, p=.006 – see Table 4.11). Here, the simple effects of level yielded significance for the female group (F(3,18)=15.542, p=.0005) but not for the male group (F(3,18)=.866, p=.477). All of the resulting pairwise comparisons between levels within the female group achieved significance other than the 55 vs. 65 dBA case. The corresponding simple effects of sex were significant for the 65 (F(1,20)=10.330, p=.004), 75 (F(1,20)=27.031, p=.0005), and 85 dBA (F(1,20)=14.598, p=.001) cases but not for the 55 dBA case (F(1,20)=3.940, p=.061, as indicated in Table 4.12.

Since no sex-level interactions were observed for the H7 and H11-H13 features, the main effects of level were calculated (Table 4.13). Both results were significant (H7: F(3,18)=9.721, p=.0005 and H11-H13:F(3,18)=13.471, p=.0005). The resulting pairwise comparisons for H7 achieved significance in the 55 vs. 85, 65 vs. 85, and 75 vs. 85 cases. All pairwise comparisons attained significance for the H11-H13 measure. Main effects of sex were also significant for both of these features (H7: F(1,20)=5.426, p=.030 and H11-H13: F(1,20)=16.430, p=.001).

The eFFRs to the 100 ms /ɔ/ vowel in Fig. 4.16 reveal a now-familiar pattern with respect to F0, H2, and H2-H6 (see Fig. 4.17). While no sex-level interactions were observed, main effects of level were significant for all three measures (F0:F(3,18)=11.739, p=.0005; H2: F(3,18)=48.903, p=.0005; H2-H6: F(3,18)=85.035, p=.0005). Here, significant differences are observed via the pairwise comparisons on F0's level means for the 55 vs. 65, 55 vs. 75, and 55 vs. 85 dBA cases. Again, the growth of H2 and that of H2-H6 appear to be robust indicators of increased sound level, with all pairwise comparisons yielding significance, and therefore exhibiting increasing growth with increasing level.

Main effects of sex were observed for the H2 (F(1,20)=7.319, p=.014) and H2-H6

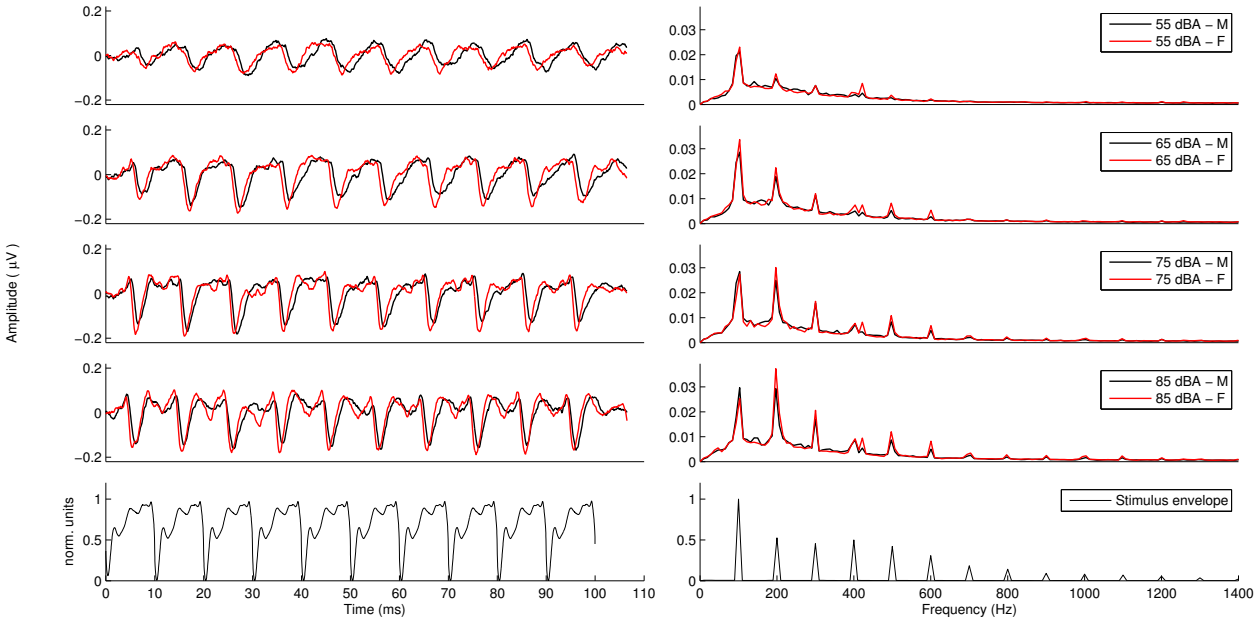

Figure 4.16: Grand-average eFFR of males (black) and females (red) to a 100 ms /ɔ/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the normalized envelope characteristics of the stimulus as calculated using the Hilbert transform. [Study 2].

Table 4.14: Statistical analysis of the eFFR to a 100ms /ɔ/ vowel for F0, H2, and H2-H6 features. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
F0	F=2.088, p=.138	**F=11.739, p=.0005**	**.662**	**.0005**	**.002**	**.010**	.025	.064	.735	F=.055, p=.818	.003
H2	F=1.322, p=.298	**F=48.903, p=.0005**	**.891**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=7.319, p=.014**	**.268**
H2-H6	F=2.319, p=.110	**F=85.035, p=.0005**	**.934**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=8.041, p=.010**	**.287**

(F(1,20)=8.041, p=.010) measures, but not for F0 (F(1,20)=.055, p=.818).

The grand-average sFFRs to the 100 ms /ɔ/ vowel shown in Fig. 4.18 also reveal a familiar pattern in that the harmonics around both F1 and F2 exhibit clear growth patterns that can be more easily observed via the plots of the mean amplitudes of low, mid, and high harmonics (Figs. 4.19 to 4.21) and are supported by the statistical test results shown in Table 4.15. Notice that there appear to be some small sex differences in the amplitude spectra, with females tending to possess larger average amplitudes than males for certain harmonic features. Despite this, no sex-level interactions were observed.

Main effects of level were observed for all four formant-related measures, however (H5-H7: F(3,18)=47.557, p=.0005; H6: F(3,18)=38.542, p=.0005; H8-H10: F(3,18)=33.998, p=.0005; H9: F(3,18)=21.248, p=.0005). All pairwise comparisons for each measure were significant, indicating consistent level-related increases in harmonic amplitude with increas-

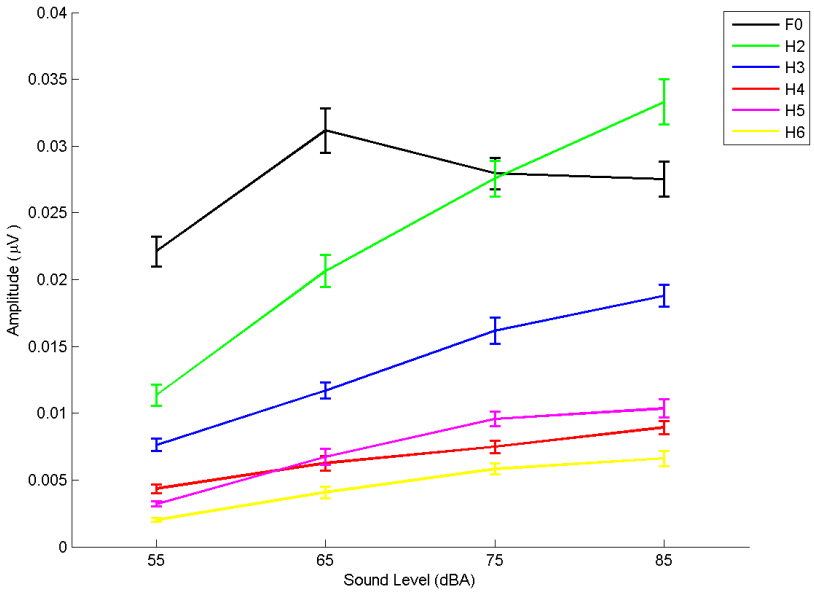

Figure 4.17: The mean eFFR amplitude spectra (in microvolts) of harmonics F0 to H6 for the 100 ms /ɔ/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

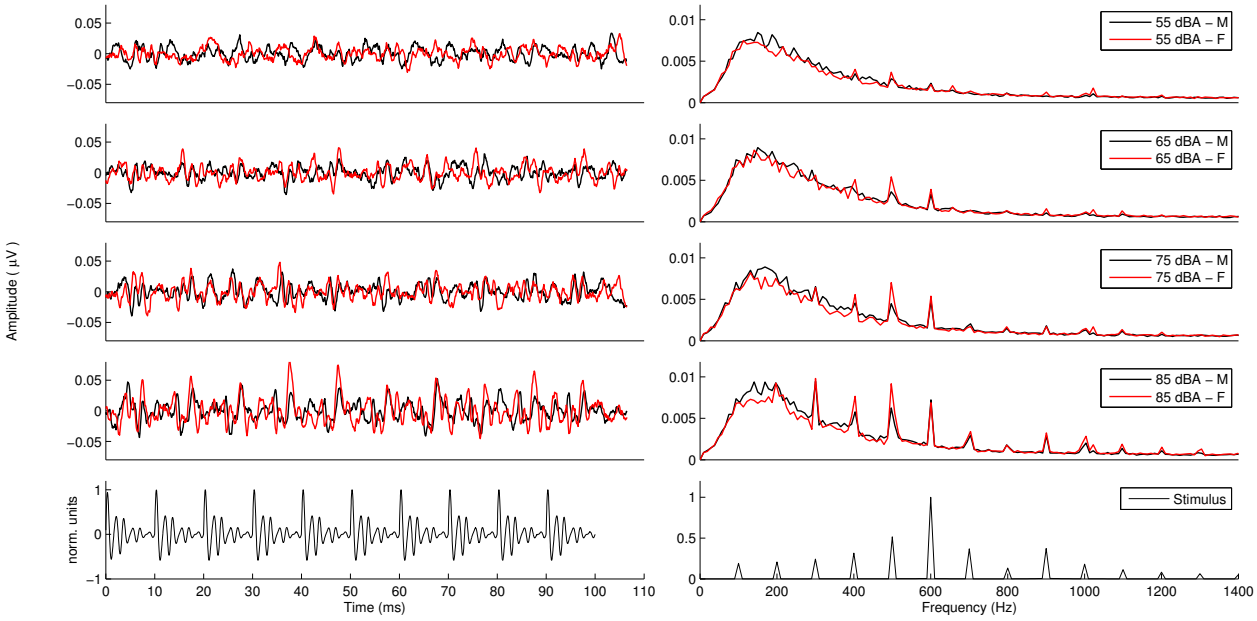

Figure 4.18: Grand-average sFFR of males (black) and females (red) to a 100 ms /ɔ/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the characteristics of the stimulus, with F1 centred at 600 Hz, and F2 centred at 900 Hz. [Study 2].

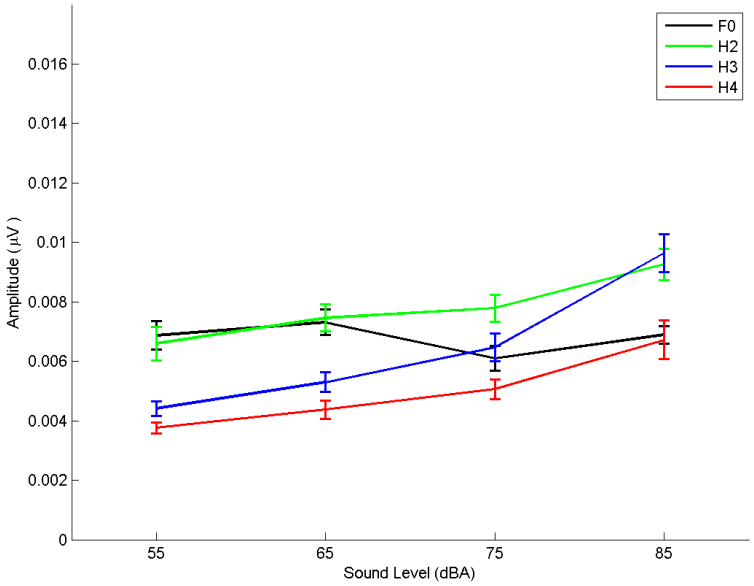

Figure 4.19: The mean sFFR amplitude spectra (in microvolts) of the low harmonics (F0 to H4) for the 100 ms /ɔ/ vowel stimulus are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

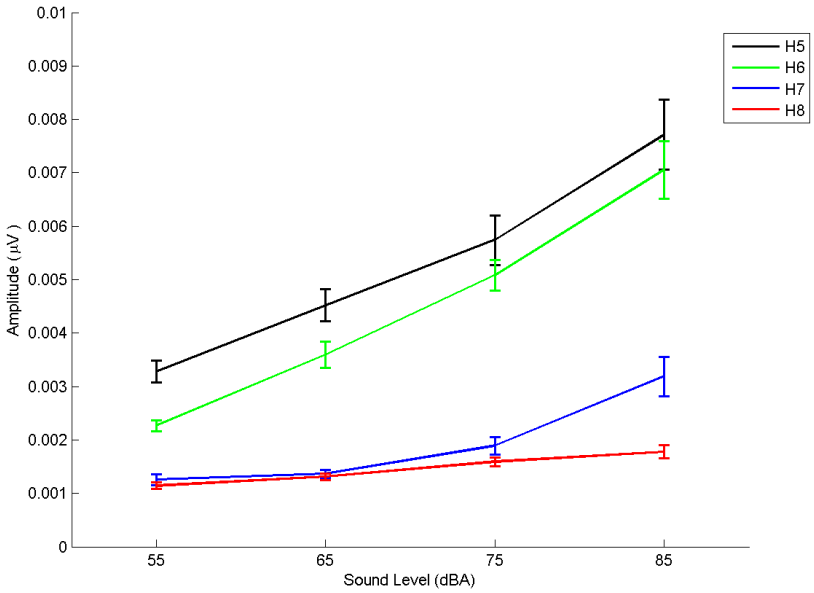

Figure 4.20: The mean sFFR amplitude spectra (in microvolts) of the mid harmonics (H5 to H8) for the 100 ms /ɔ/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

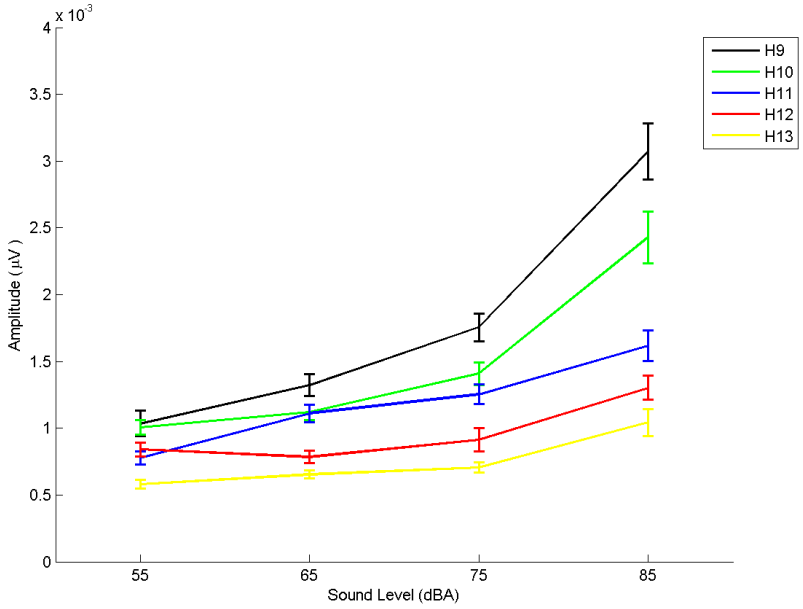

Figure 4.21: The mean sFFR amplitude spectra (in microvolts) of the high harmonics (H9 to H13) for the 100 ms /ɔ/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

ing level. Main effects of sex were also observed for the H5-H7 (F(1,20)=9.878, p=.005), H8-10 (F(1,20)=9.098, p=.007), and H9 (F(1,20)=4.754, p=.041) measures, but not for the H6 measure (F(1,20)=.146, p=.706).

Although the grand-average eFFRs to the 100 ms /U/ vowel stimulus shown in Fig. 4.22 appear to follow a similar pattern of the previously examined stimuli, the mean amplitudes of the harmonic spectra from H2 to H6 generally exhibited only slight growth from 65 to 85 dBA (see Fig. 4.23). F0 once again exhibits a pattern of growth, saturation and decline here. These visual observations were confirmed via statistical testing (see Tables 4.16 to 4.18). A sex-level interaction was observed for the F0 measure (F(3,18)=3.766, p=.029) with both

Table 4.15: Statistical analysis of the sFFR to a 100ms /ɔ/ vowel. The F1 (H6) and F2 (H9) centre harmonics were tested for increasing growth with increasing level, as well as their more global formant measures via H5-H7 and H8-H10 combinations, respectively. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
H5-H7	F=1.567, p=.232	**F=47.557, p=.0005**	**.888**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=9.878, p=.005**	**.331**
H6	F=1.960, p=.156	**F=38.542, p=.0005**	**.865**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	F=.146, p=.706	.007
H8-H10	F=.6445, p=.596	**F=33.998, p=.0005**	**.850**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=9.098, p=.007**	**.313**
H9	F=2.676, p=.078	**F=21.248, p=.0005**	**.780**	**.0015**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=4.754, p=.041**	**.192**

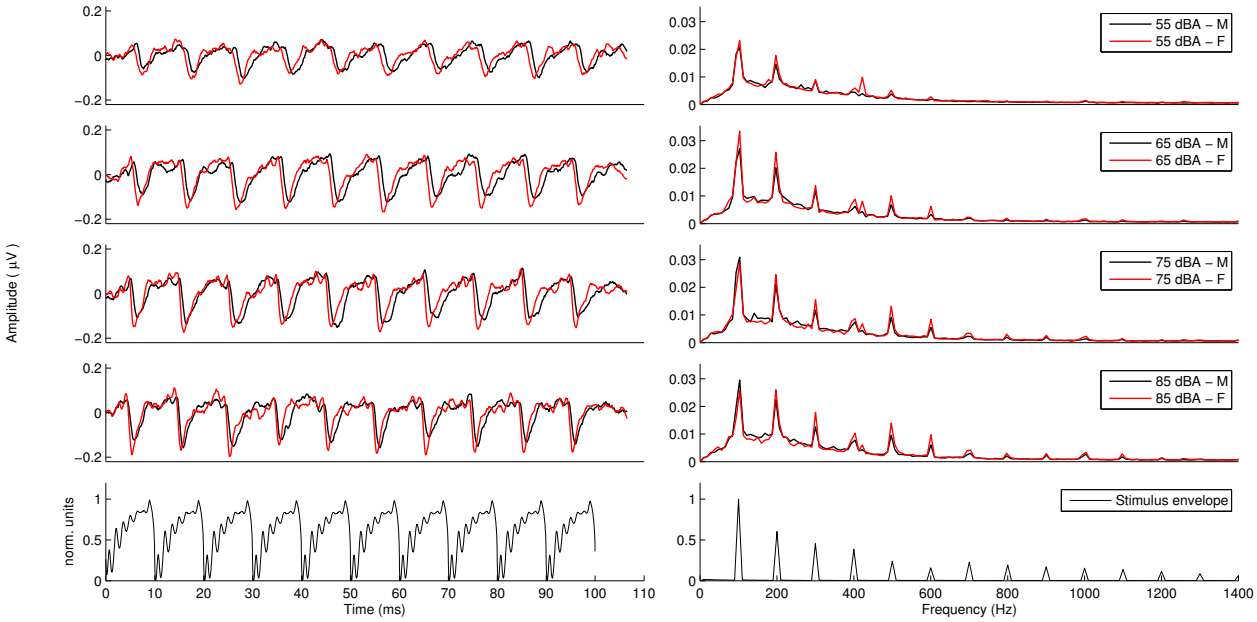

Figure 4.22: Grand-average eFFR (red) to a 100 ms /U/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the normalized envelope characteristics of the stimulus as calculated using the Hilbert transform. [Study 2].

Table 4.16: Statistical analysis of the eFFR to a 100 ms /U/ vowel on the F0, H2, and H2-H6 features. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Simple effect of level**		**Pairwise comparisons between sound levels (p-values)**					
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
F0 (F)	**F=3.766, p=.029**	**F=12.756, p=.0005**	**.680**	**.0005**	.020	.377	.045	**.003**	.227
F0 (M)	**F=3.766, p=.029**	**F=.459, p=.002**	**.541**	**.003**	**.0005**	**.006**	.098	.312	.571

the female (F(3,18)=12.756, p=.0005) and male F(3,18)=.459, p=.002) tests of simple effects of level attaining significance. The pairwise comparisons between levels revealed significant differences on the 55 vs. 65 and 65 vs. 85 dBA tests. For the females, significant differences were observed for the 55 vs. 65, 55 vs. 75, and 55 vs. 85 dBA comparisons. No simple effects of sex were observed here for F0 (see Table 4.17).

Since no sex-level interactions were observed for the H2 and H2-H6 measures, main effects of level were tested, and both results bore significance (H2: F(3,18)=13.395, p=.0005; H2-

Table 4.17: Simple effects of sex within each level condition of F0 for the eFFR to a 100 ms /U/ vowel (corresponding to the sex-level interactions noted in Table 4.16). [Study 2].

	55 dBA		**65 dBA**		**75 dBA**		**85 dBA**	
Feat.	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2	**F(1,20), p-val.**	η_p^2
F0	F=.664, p=.425	.032	F=2.758, p=.112	.121	F=.716, .408	.035	F=1.155, p=.295	.055

Table 4.18: Main effects results for the non-significant sex-level interactions of the eFFR to a 100 ms /U/ vowel. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F, p-val.**	η_p^2
H2	F=.392, p=.760	**F=13.395, p=.0005**	**.691**	**.0005**	**.0005**	**.0005**	.546	.2495	.207	**F=7.893, p=.010**	**.285**
H2-H6	F=1.328, p=.296	**F=31.930, p=.0005**	**.842**	**.0005**	**.0005**	**.0005**	.0985	**.0165**	.033	**F=13.640, p=.001**	**.405**

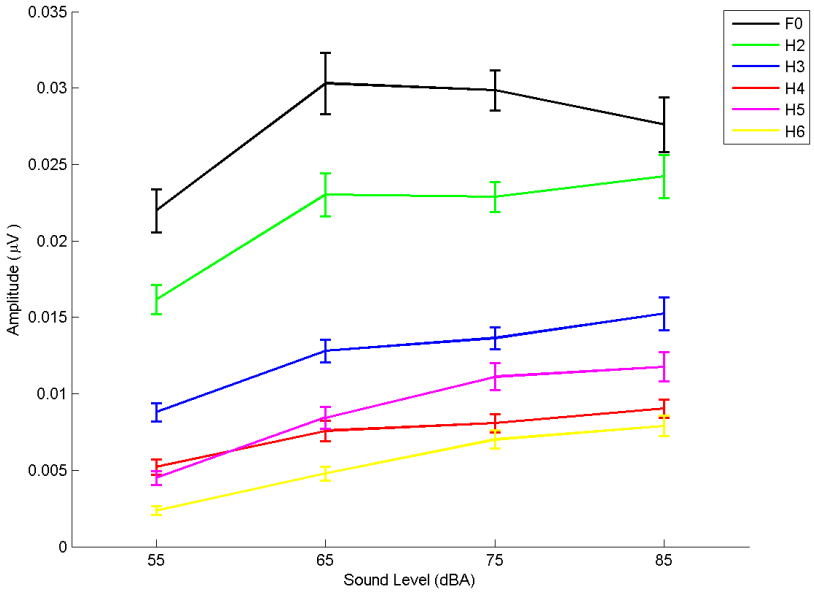

Figure 4.23: The mean eFFR amplitude spectra (in microvolts) of harmonics F0 to H6 for the 100 ms /U/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

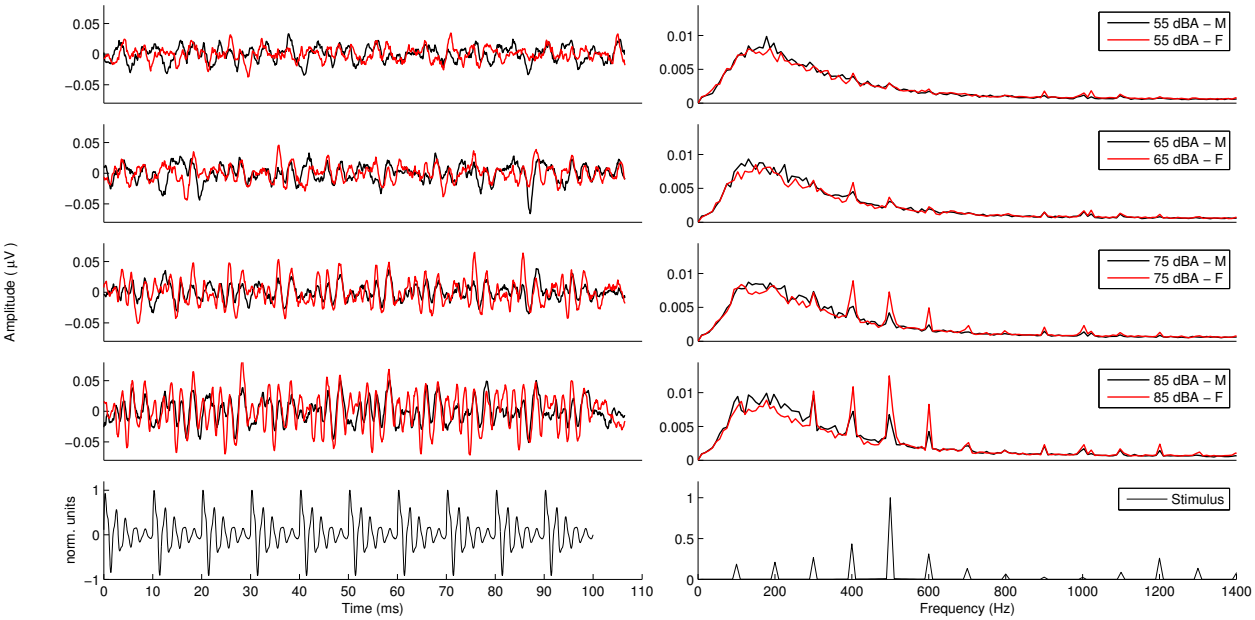

Figure 4.24: Grand-average sFFR of males (black) and females (red) to a 100 ms /U/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the characteristics of the stimulus, with F1 centred at 500 Hz, and F2 centred at 1200 Hz. [Study 2].

H6: F(3,18)=31.930, p=.0005). The pairwise comparisons for both revealed mean differences between the 55 vs. 65, 55 vs. 75, and 55 vs. 85 dBA cases, with the H2-H6 measure additionally attaining significance on the 65 vs. 85 dBA comparison. This is consistent with the growth pattern observed visually in Fig. 4.23. Main effects of sex were also observed for these two measures (H2: F(1,20)=7.893, p=.010; and H2-H6: F(1,20)=13.640, p=.001).

The grand-average of the sFFRs to the 100 ms /U/ vowel stimulus shown in Fig. 4.24 once again follows the familiar pattern of increasing amplitude spectra in the formant regions with increasing level. Growth patterns for the low, mid and high harmonic ranges are displayed in Figs. 4.25 to 4.27. Despite the visible sex differences between mean amplitude spectra – particularly apparent from H4-H6 – no sex-level interactions were observed here (see Table 4.19). Main effects of level for all of the formant related measures were significant (H4-H6: F(3,18)=17.022, p=.0005; H5: F(3,18)=17.057, p=.0005; H11-H13: F(3,18)=26.779, p=.0005). The global F1 measure (H4-H6 here) attains significance amongst all pairwise comparisons. The follow-up tests on the H5 level-related means on all but the 55 vs. 65 dBA comparisons. The second formant follow-up tests on the H11-H13 and H12 measures achieve significance on the 55 vs. 85, 65 vs. 85, and 75 vs. 85 dBA comparisons. Additionally, the H11-H13 measure attains significance on the 55 vs. 65 dBA test, and the H12 measures attains significance on the 55 vs. 75 dBA test. Main effects of sex were observed for the H11-H13 (F(1,20)=14.166, p=.001) and H12 measures (F(1,20)=18.719, p=.0005).

The grand-average eFFRs to the 100 ms /u/ vowel shown in Fig. 4.28 are atypical in

Table 4.19: Statistical analysis of the sFFR to a 100 ms /U/ vowel. The F1 (H5) and F2 (H12) centre harmonics were tested for increasing growth with increasing level, as were their global formant features H4-H6 and H11-H13, respectively. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
H4-H6	F=2.208, p=.146	**F=17.022, p=.0005**	**.739**	**.048**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	F=3.504, p=.076	.149
H5	F=2.149, p=.130	**F=17.057, p=.0005**	**.740**	.070	**.0015**	**.0005**	**.001**	**.0005**	**.0005**	F=3.719, p=.068	.157
H11-H13	F=1.713, p=.200	**F=26.779, p=.0005**	**.817**	**.004**	.947	**.0005**	.8445	**.0005**	**.0005**	**F=14.166, p=.001**	**.415**
H12	F=1.499, p=.249	**F=13.430, p=.0005**	**.691**	.0835	**.0055**	**.0005**	.143	**.0005**	**.0005**	**F=18.719, p=.0005**	**.483**

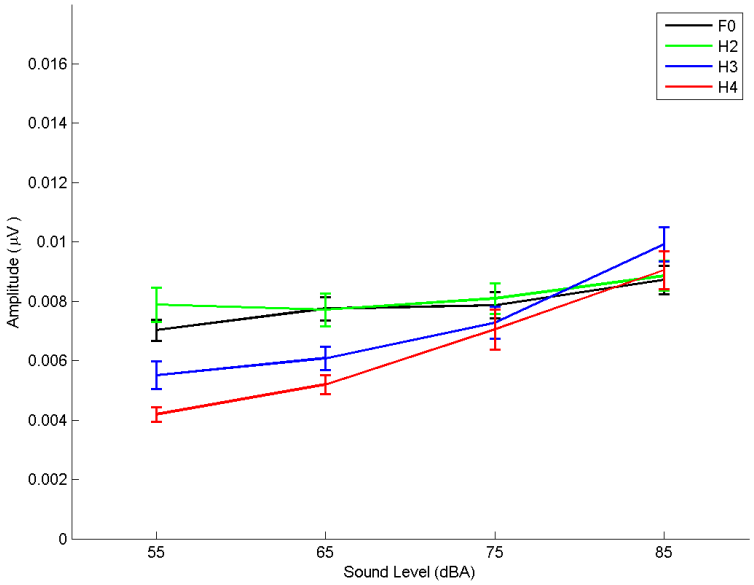

Figure 4.25: The mean sFFR amplitude spectra (in microvolts) of the low harmonics (F0 to H4) for the 100 ms /U/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

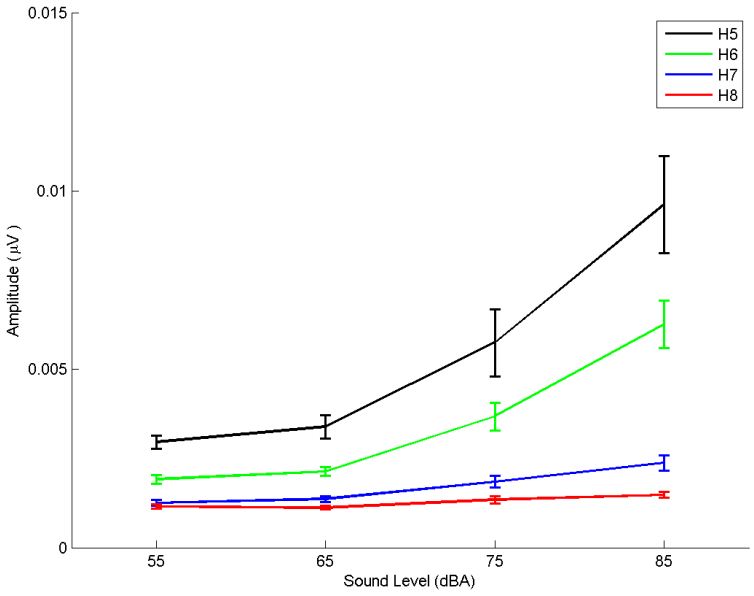

Figure 4.26: The mean sFFR amplitude spectra (in microvolts) of the mid harmonics (H5 to H8) for the 100 ms /U/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

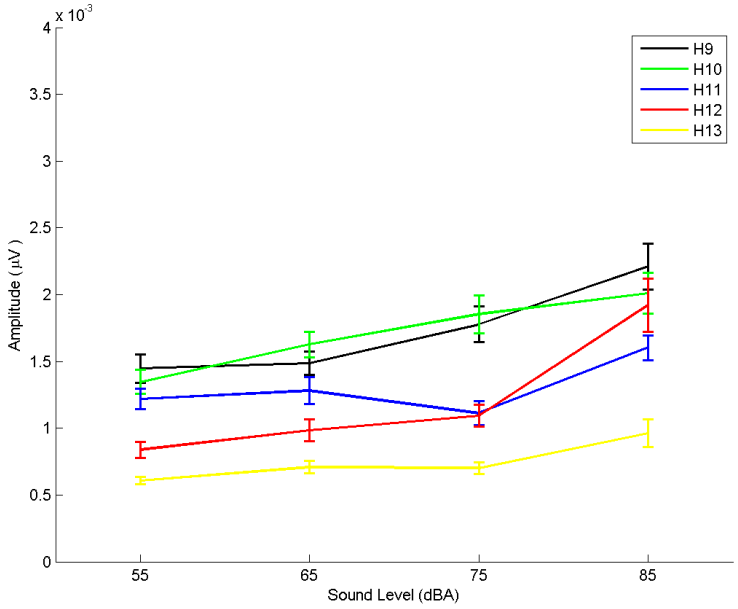

Figure 4.27: The mean sFFR amplitude spectra (in microvolts) of the high harmonics (H9 to H12) for the 100 ms /U/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

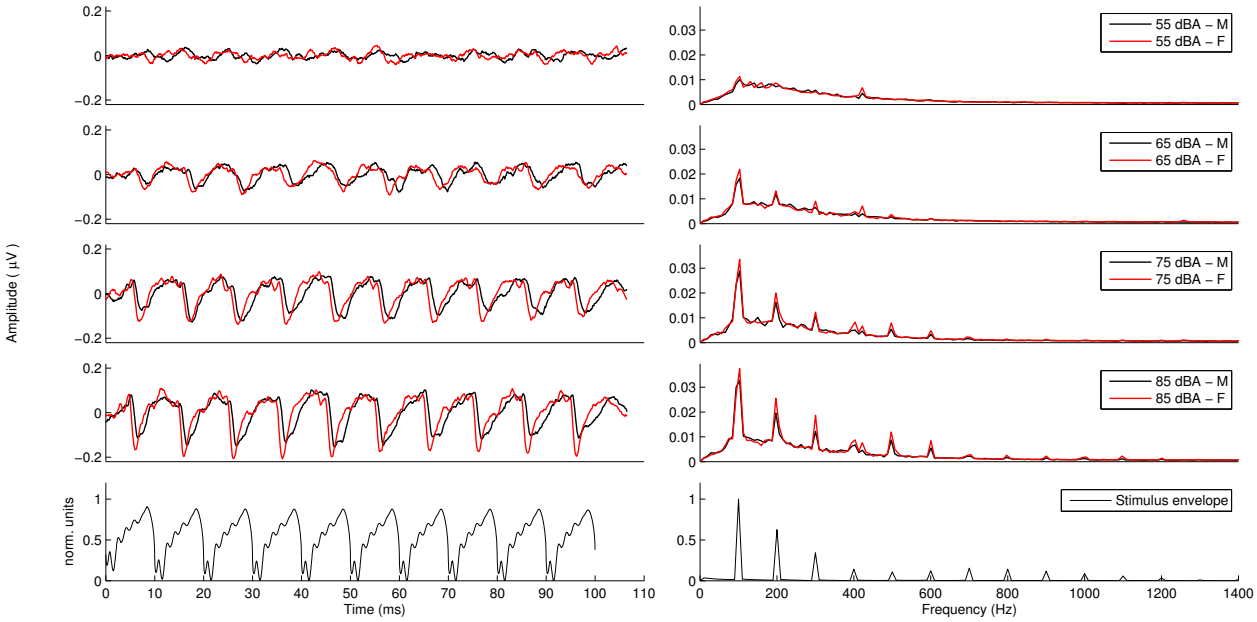

Figure 4.28: Grand-average eFFR of males (black) and females (red) to a 100 ms /u/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the normalized envelope characteristics of the stimulus as calculated using the Hilbert transform. [Study 2].

Table 4.20: Statistical analysis of the eFFR to a 100ms /u/ vowel for F0, H2, and H2-H6 features. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
F0	F=.785, p=.518	**F=86.602, p=.0005**	**.932**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0055**	F=2.521, p=.128	.112
H2	F=.638, p=.600	**F=30.461, p=.0005**	**.835**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.002**	**F=4.801, p=.040**	**.194**
H2-H6	F=2.652, p=.080	**F=58.969, p=.0005**	**.908**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**F=10.139, p=.005**	**.336**

comparison to what has been observed thus far. There is a very clear monotonic increasing growth of F0 and H2 is relatively small in amplitude compared to F0 here when contrasted with the other stimuli studied. These trends and the growth functions of the remaining harmonics under study can be seen more clearly in Fig. 4.29. Although some minor sex-related differences are visible, no sex-level interactions were observed (see Table 4.20 for a summary of the statistical results here). Main effects of level were observed for all three measures of F0 (F(3,18)=86.602, p=.0005), H2 (F(3,18)=30.461, p=.0005), and H2-H6 (F(3,18)=58.969, p=.0005), and all pairwise comparisons for each measure attained significance as well. Main effects of sex were observed for H2 (F(1,20)=4.801, p=.040) and H2-H6 (F(1,20)=10.139), p=.005, but not for F0 (F(1,20)=2.521, p=.128).

The grand-average sFFRs to the 100 ms /u/ vowel stimulus shown in Fig. 4.30 are consistent with the other results presented, although the formant growth appears relatively

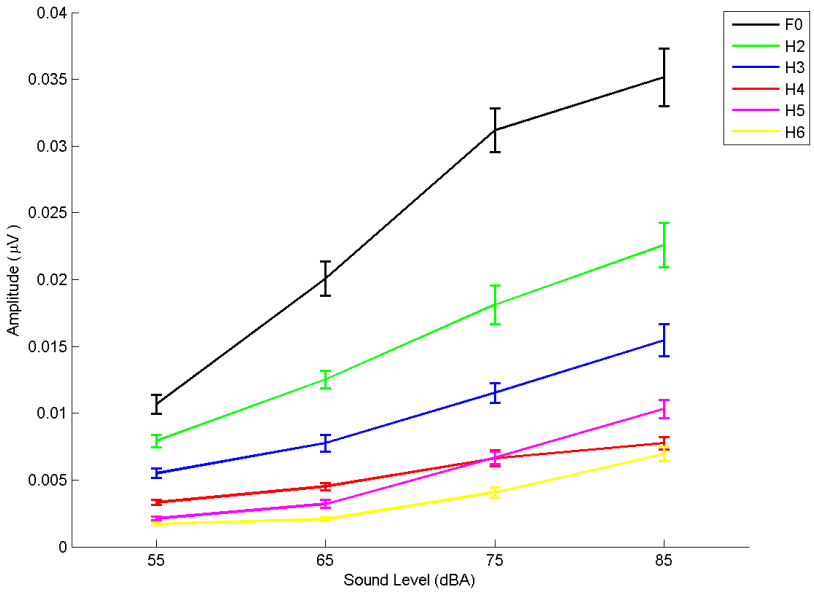

Figure 4.29: The mean eFFR amplitude spectra (in microvolts) of harmonics F0-H6 for the 100 ms /u/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

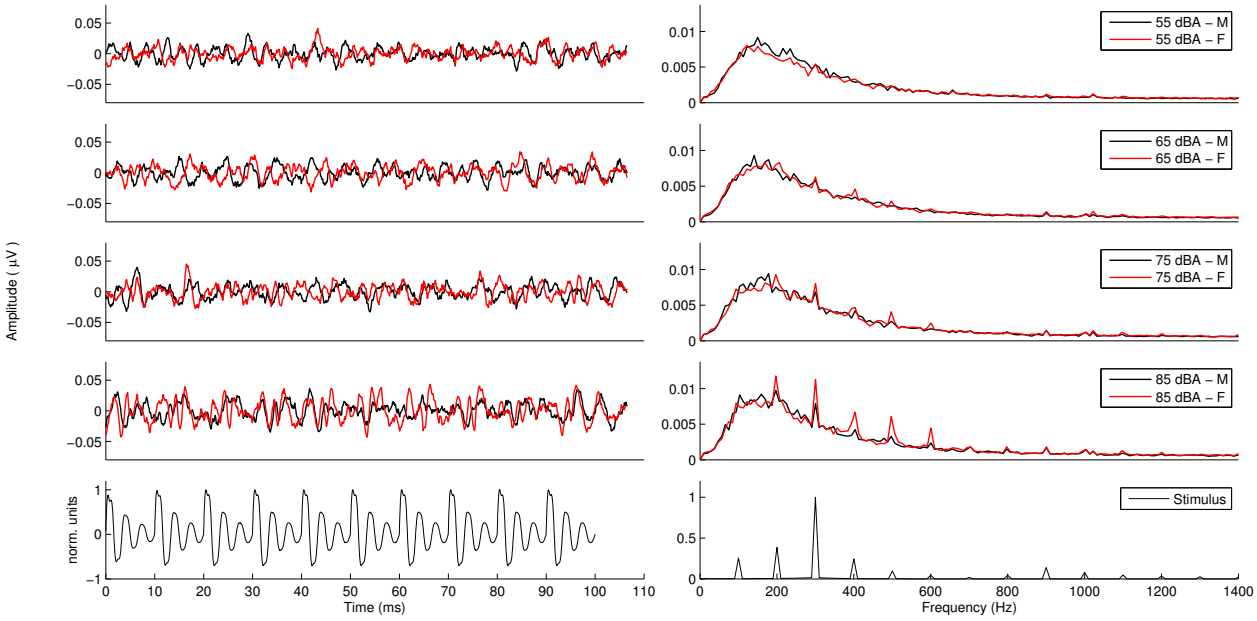

Figure 4.30: Grand-average sFFR of males (black) and females (red) to a 100 ms /u/ vowel with F0=100 Hz presented at 55, 65, 75, and 85 dBA. Both the waveform (left), and amplitude spectra (right) representations are shown. The bottom panel shows the characteristics of the stimulus, with F1 centred at 300 Hz, and F2 centred at 900 Hz. [Study 2].

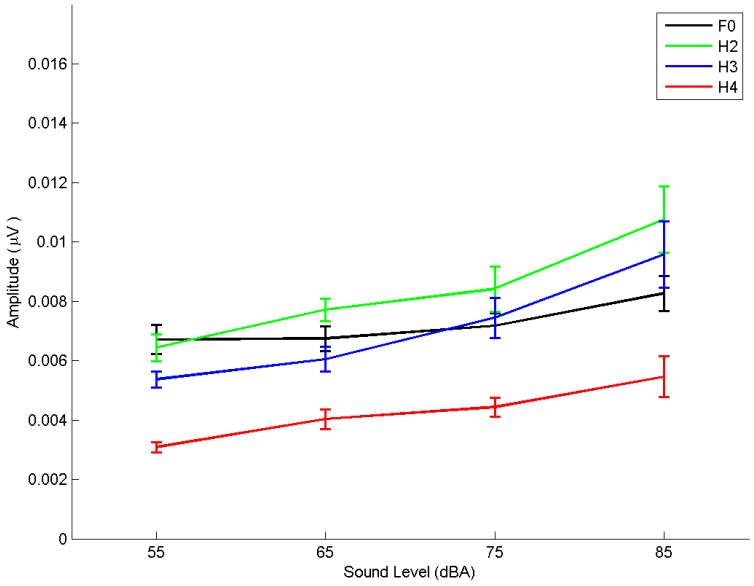

Figure 4.31: The mean sFFR amplitude spectra (in microvolts) of low harmonics (F0 to H4) for the 100 ms /u/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

weaker than previously observed. The growth functions of the low, mid, and high harmonics (Figs. 4.31 to 4.33) tend to exhibit monotonic increasing behaviour throughout. Despite some visible sex differences in the harmonic spectra, no sex-level interactions were observed (see Table 4.21 for a summary of the statistical tests here). Main effects of level were observed for all four formant-related measures (H2-H3: $F(3,18)=8.246$, $p=.001$; H4: $F(3,18)=4.900$, $p=.012$; H8-H10: $F(3,18)=20.301$, $p=.0005$; H9: $F(3,18)=7.009$, $p=.003$). All pairwise comparisons attained significance for the H2-H4 measure, as did all but the 55 vs. 65 dBA test for the H3 measure. All of the H8-H10 pairwise comparisons also attained significance, whereas the H9 measure's comparisons were only significant for the 55 vs. 65, 55 vs. 75, 55 vs. 85, and 65 vs. 85 dBA tests. No main effects of sex were observed.

Time domain analysis

The grand-average eFFR and sFFR waveforms for each of the 100 ms /a/, /ɔ/, /U/, and /u/ vowels are displayed in Figs. 4.10, 4.16, 4.22, 4.28 and Figs. 4.12, 4.18, 4.24, and 4.30, respectively. The mean interpeak intervals for each vowel are tabulated at each sound level (Table 4.22) and once again show a very precise tracking of the pitch period of 10 ms in all cases (1/F0).

Figure 4.34 displays the relative normalized latency of the 100 ms /a/ vowel's ten major waveform peaks at each sound level. Clearly, there is a systematic shift in latency, with

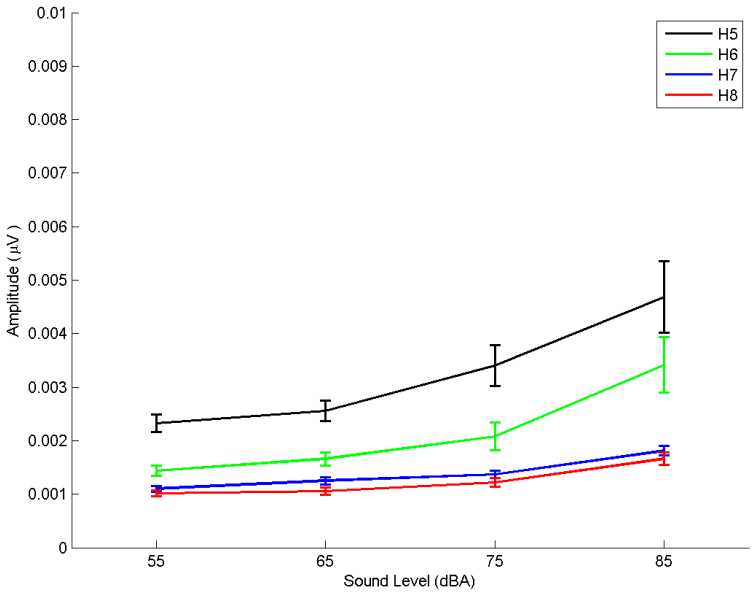

Figure 4.32: The mean sFFR amplitude spectra (in microvolts) of low harmonics (H5 to H8) for the 100 ms /u/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

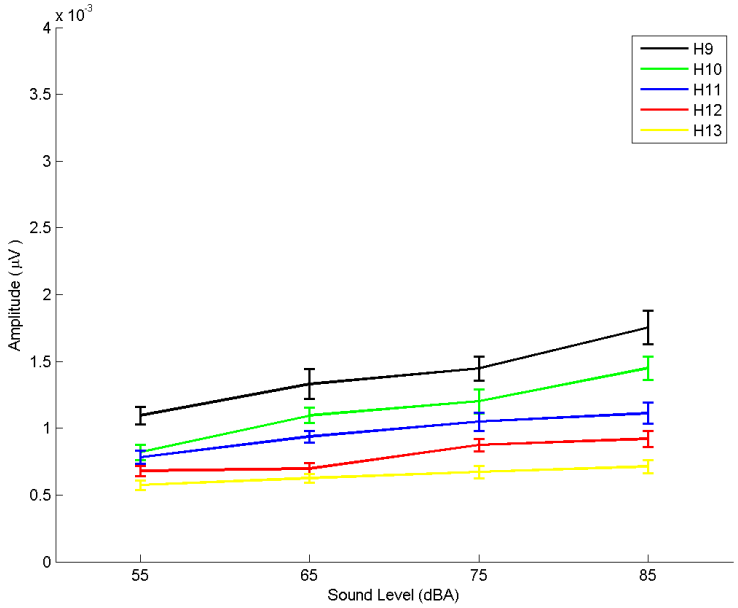

Figure 4.33: The mean sFFR amplitude spectra (in microvolts) of high harmonics (H9 to H12) for the 100 ms /u/ vowel are plotted as a function of sound level. The error bars indicate the SEM. [Study 2].

Table 4.21: Statistical analysis of the sFFR to a 100ms /u/ vowel. The F1 (H3) and F2 (H9) centre harmonics were tested for increasing growth with increasing level, as were their global formant features H2-H4 and H8-H10, respectively. Bold numbers indicate statistically significant results. [Study 2].

	Inter. effect	**Main effect of level**		**Pairwise comparisons between sound levels (p-values)**						**Main effect of sex**	
Feat.	**F(3,18), p-val.**	**F(3,18), p-val.**	η_p^2	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**	**F(1,20), p-val.**	η_p^2
H2-H4	F=.819, p=.500	**F=8.246, p=.001**	**.579**	**.0035**	**.0015**	**.0005**	**.0405**	**.001**	**.0005**	F=1.448, p=.243	.068
H3	F=.863, p=.478	**F=4.900, p=.012**	**.450**	.0875	**.002**	**.0005**	**.0145**	**.001**	**.002**	F=1.585, p=.223	.073
H8-H10	F=.543, p=.659	**F=20.301, p=.0005**	**.772**	**.001**	**.0005**	**.0005**	**.022**	**.0005**	**.0005**	F=1.089, p=.309	.052
H9	F=.660, p=.587	**F=7.009, p=.003**	**.539**	**.016**	**.0005**	**.0005**	.1095	**.0075**	.0255	F=.685, p=.418	.033

Table 4.22: Mean (SD) interpeak intervals (ms) for all four 100 ms vowel stimuli at each sound level. [Study 2].

	Mean interpeak interval (ms) per stimulus			
Level (dBA)	**/a/ vowel**	**/ɔ/ vowel**	**/U/ vowel**	**/u/ vowel**
55	10.0 (.6)	10.0 (.7)	10.0 (.7)	10.0 (1.1)
65	10.0 (.4)	10.0 (.5)	10.0 (.4)	10.0 (.8)
75	10.0 (.4)	10.0 (.4)	10.0 (.4)	10.0 (.5)
85	10.0 (.2)	10.0 (.3)	10.0 (.3)	10.0 (.5)

higher levels producing shorter latencies, as expected.

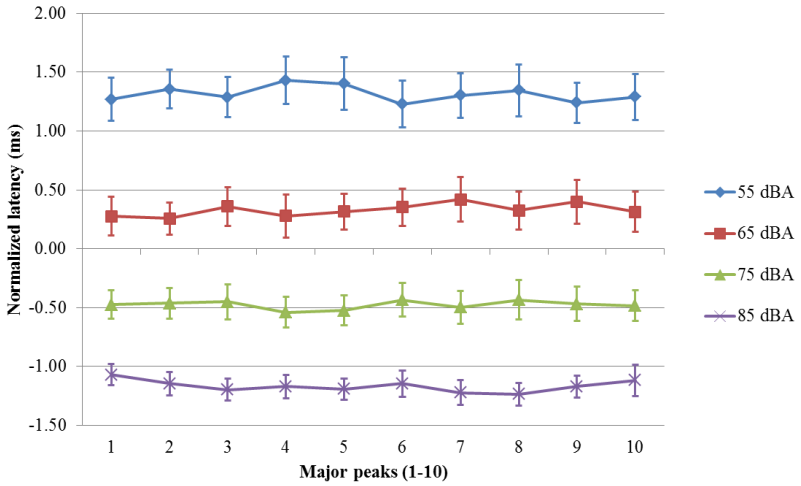

Figure 4.34: Relative (normalized) latency of the major peaks present in the eFFRs to a 100 ms /a/ vowel as a function of sound level. The more positive the latency value, the later the peak occurred and vice-versa. The error bars represent the standard error of the mean. The trend clearly indicates that the greater the sound level, the shorter the latency of the peak. [Study 2].

A mixed-effect RMANOVA on the normalized means of the 100 ms /a/ vowel at each sound level revealed no significant sex-level interaction (p=.115); however, there was a significant main effect of sound level (F(3,18)=143.819, p<.0005, η_p^2 = .960). All pairwise

Table 4.23: Results of RMANOVAs on the latencies of each peak of the 100 ms /a/ vowel as a function of sound level (see Fig 4.34), and the subsequent pairwise comparisons performed between levels. Note that all tests obtain statistical significance (indicated in bold). [Study 2].

	Main effect of level		**Pairwise comparisons between sound levels (p-values)**					
Peak	**F(3,19)**	**p-val.**	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
1	**64.861**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
2	**108.291**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
3	**97.690**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
4	**78.688**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
5	**74.075**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
6	**46.577**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
7	**82.167**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
8	**78.129**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
9	**87.977**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
10	**59.301**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**

comparisons between level means were significant with p<.0005. The ten major peaks were therefore tested for level effects individually using RMANOVAs and follow-up one-sided t-tests to compare means between each pair of conditions. All said tests attained significance, and are summarized in Table 4.23.

The mixed-effect RMANOVA also revealed a significant main effect of sex (F(1,20)=17.420, p<0.0005, η_p^2=.466). The pairwise comparison between sexes revealed females to have shorter latencies than males (p<0.0005) by an average of .79 ms.

The corresponding amplitude of each major peak is shown in Fig. 4.35. Based on piloting, and the results of Study 1, this measure was not generally expected to be a robust indicator of level change and so no statistical tests were planned. It is worth noting that the amplitudes are determined here using detrended signals and so the absolute deviation away from zero reveals that the amplitudes of the 55 dBA case tend to be smaller than for the other three conditions.

The relative normalized latencies of the 10 major negative peaks in the eFFR waveforms of the 100 ms /ɔ/ vowel are shown in Fig. 4.36. They follow the same trend seen previously, with lower levels resulting in later peaks, and vice-versa.

A mixed-effect RMANOVA on the normalized means at each sound level revealed no significant sex-level interaction (p=.504), however there was a significant main effect of sound level (F(3,18)=108.500, p<.0005, η_p^2 =.948). All pairwise comparisons between level means were significant with p<.0005. The ten major peaks were therefore tested individually for level effects using RMANOVAs and follow-up one-sided t-tests to compare means between each pair of conditions. The results of the consequent RMANOVAs and their associated pairwise comparisons, all of which were significant, are shown in Table 4.24.

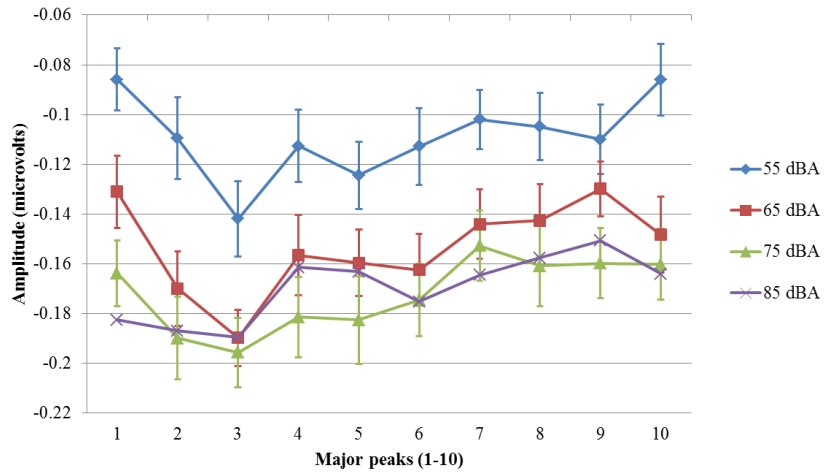

Figure 4.35: Major peak amplitudes as a function of sound level for the 100 ms /a/ vowel stimulus. Since the most obvious stable peaks across all stimuli were the negative peaks, and the subaverages were all detrended, the absolute value of the amplitudes provides a good sense of the relative size of the voltage fluctuation differences between conditions. Therefore, the 55 dBA case has the smallest peaks relative to the three other level conditions, which are all fairly similar. [Study 2].

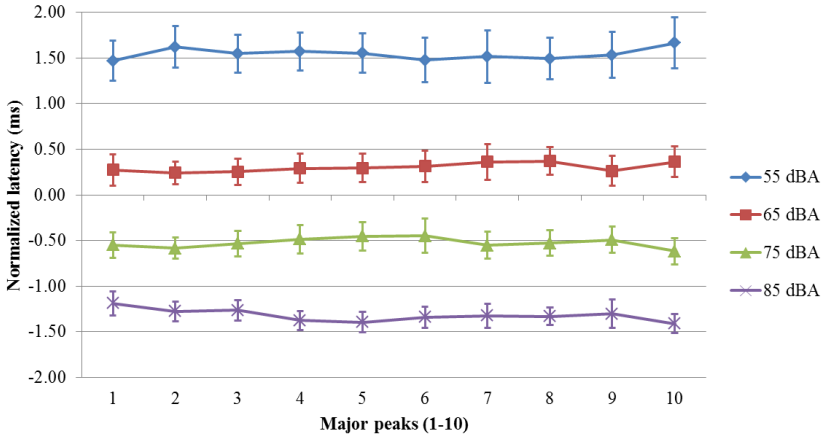

Figure 4.36: Relative (normalized) latency of the major peaks present in the eFFRs to a 100 ms /ɔ/ vowel as a function of sound level. The more positive the latency value, the later the peak occurred and vice-versa. The error bars represent the standard error of the mean. The trend clearly indicates that the greater the sound level, the shorter the latency of the peak. [Study 2].

The mixed-effect RMANOVA also revealed a significant main effect of sex ($F(1,20)=7.319$, $p<0.014$, $\eta_p^2=.268$). The pairwise comparison between sexes revealed females to have shorter latencies than males ($p<0.014$) by an average of .64 ms.

The amplitudes corresponding to the same major peaks in Fig. 4.36 are shown in Fig. 4.37. They are similar to the results for the 100 ms /a/ vowel, although here the 65 dBA case clearly overlaps less with the higher level conditions and appears to be nearly separable from the 85 dBA condition.

The relative normalized latencies of the 10 major negative peaks in the eFFR of the

Table 4.24: Results of RMANOVAs on the latencies of each peak of the 100 ms /ɔ/ vowel (see Fig. 4.36) as a function of sound level, and the subsequent pairwise comparisons performed between levels. Note that all tests obtain statistical significance (indicated in bold). [Study 2].

	Main effect of level		Pairwise comparisons between sound levels (p-values)					
Peak	**F(3,19)**	**p-val.**	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
1	**73.268**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
2	**67.839**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
3	**89.066**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
4	**109.102**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
5	**103.973**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
6	**74.206**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
7	**59.456**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
8	**85.473**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
9	**64.916**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
10	**63.181**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**

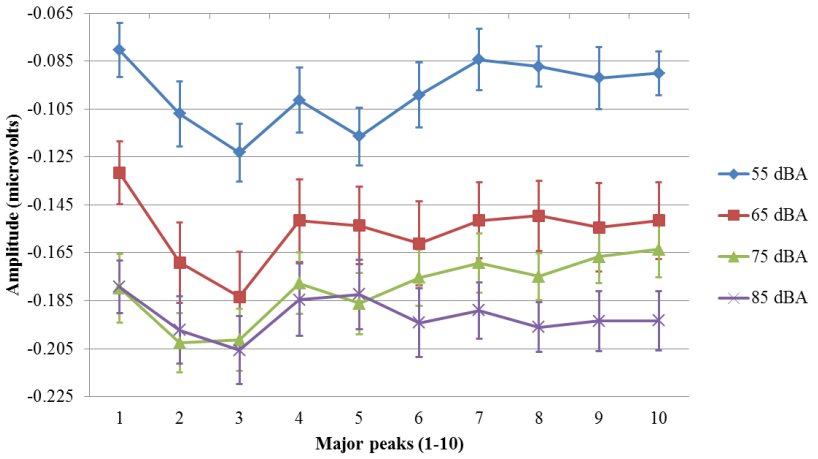

Figure 4.37: Major peak amplitudes as a function of sound level for the 100 ms /ɔ/ vowel stimulus. [Study 2].

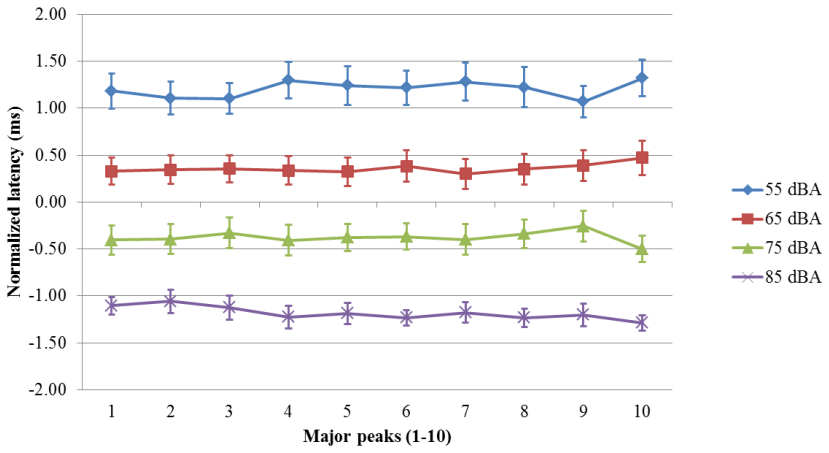

Figure 4.38: Relative (normalized) latency of the major peaks present in the eFFRs to a 100 ms /U/ vowel as a function of sound level. The more positive the latency value, the later the peak occurred and vice-versa. The error bars represent the standard error of the mean. The trend clearly indicates that the greater the sound level, the shorter the latency of the peak. [Study 2].

100 ms /U/ vowel are shown in Fig. 4.39. Again, they follow the same trend as seen in the analysis of the /a/ and /ɔ/ vowels, wherein the lower levels result in later peaks, and vice-versa.

A mixed-effect RMANOVA on the normalized means at each sound level revealed no significant sex-level interaction (p=0.339), however there was a significant effect of sound level ($F(3,18)=156.636$, $p<.0005$, $\eta_p^2 = .963$). All pairwise comparisons between level means were significant with $p<0.0005$. The ten major peaks were therefore tested for level effects individually using RMANOVAs and follow-up one-sided t-tests to compare means between each pair of conditions. The results of the consequent RMANOVAs and their associated pairwise comparisons, all of which were significant, are shown in Table 4.25.

The mixed-effect RMANOVA also revealed a significant main effect of sex ($F(1,20)=18.464$, $p<0.0005$, $\eta_p^2=.480$). The pairwise comparison between sexes revealed females to have shorter latencies than males ($p<0.0005$) by an average of .8 ms.

The amplitudes corresponding to the major peaks in Fig. 4.38 are shown in Fig. 4.39. They are similar to those of the 100 ms /a/ vowel shown in Fig. 4.35 in that the 55 dBA condition has markedly smaller absolute amplitudes than the remaining three conditions, which overlap substantially with one another.

Finally, the relative normalized latencies of the 10 major negative peaks in the eFFR of the 100 ms /u/ vowel are shown in Fig. 4.41. They are similar to those of the previous three stimuli; however there is a noticeably larger difference between the normalized latencies of the 55 and 65 dBA peaks here. This is consistent with the observation that the amplitude spectra at 55 dBA are relatively smaller than those of the other three vowel responses (see

Table 4.25: Results of RMANOVAs on the latencies of each peak of the 100 ms /U/ vowel (see Fig. 4.38) as a function of sound level, and the subsequent pairwise comparisons performed between levels. Note that all tests obtain statistical significance (indicated in bold). [Study 2].

	Main effect of level		**Pairwise comparisons between sound levels (p-values)**					
Peak	**F(3,19)**	**p-val.**	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
1	**115.125**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
2	**94.048**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
3	**94.602**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
4	**91.516**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
5	**82.849**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
6	**78.785**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
7	**62.518**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
8	**71.983**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
9	**55.358**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
10	**62.488**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**

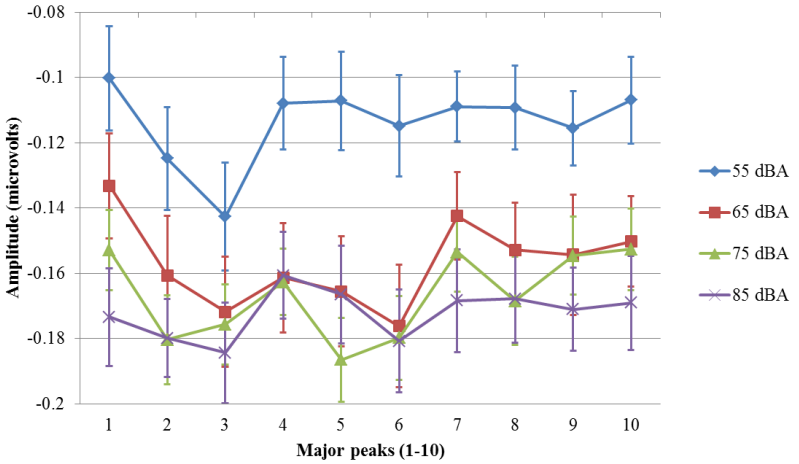

Figure 4.39: Major peak amplitudes as a function of sound level for the 100 ms /U/ vowel stimulus. [Study 2].

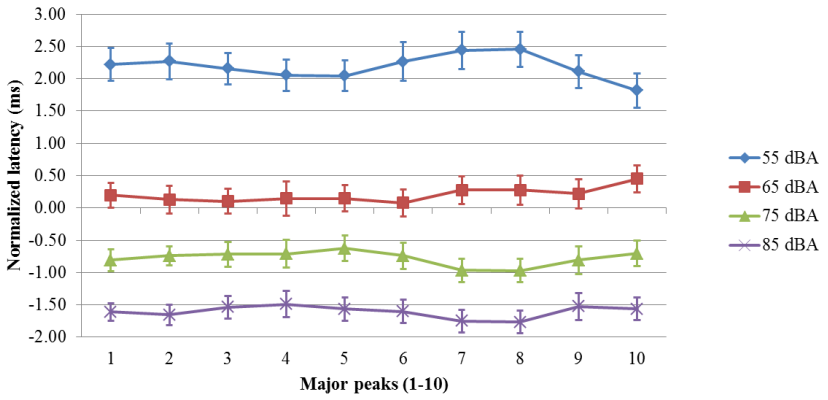

Figure 4.40: Relative (normalized) latency of the major peaks present in the eFFRs to a 300 ms /u/ vowel as a function of sound level. The more positive the latency value, the later the peak occurred and vice-versa. The error bars represent the standard error of the mean. The trend clearly indicates that the greater the sound level, the shorter the latency of the peak. [Study 2].

Table 4.26: Results of RMANOVAs on the latencies of each peak of the 100 ms /u/ vowel (see Fig. 4.40) as a function of sound level, and the subsequent pairwise comparisons performed between levels. Note that all tests obtain statistical significance (indicated in bold). [Study 2].

	Main effect of level		**Pairwise comparisons between sound levels (p-values)**					
Peak	**F(3,19)**	**p-val.**	**55 vs 65**	**55 vs 75**	**55 vs 85**	**65 vs 75**	**65 vs 85**	**75 vs 85**
1	**253.111**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
2	**76.990**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
3	**108.424**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
4	**97.532**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
5	**100.041**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
6	**114.898**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
7	**104.721**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
8	**180.401**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
9	**66.767**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**
10	**91.236**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**	**.0005**

Fig. 4.29).

A mixed-effect RMANOVA on the normalized means at each sound level revealed no significant sex-level interaction (p=0.922), however there was a significant effect of sound level ($F(3,18)=248.012$, $p<.0005$, $\eta_p^2=0.976$). All pairwise comparisons between level means were significant with $p<0.0005$. The ten major peaks were therefore tested for level effects individually using RMANOVAs and follow-up one-sided t-tests to compare means between each pair of conditions. The results of the consequent RMANOVAs and their associated pairwise comparisons, all of which were significant, are shown in Table 4.26.

The mixed-effect RMANOVA also revealed a significant main effect of sex ($F(1,20)=19.221$, $p<0.0005$, $\eta_p^2=.490$). The pairwise comparison between sexes revealed females to have shorter

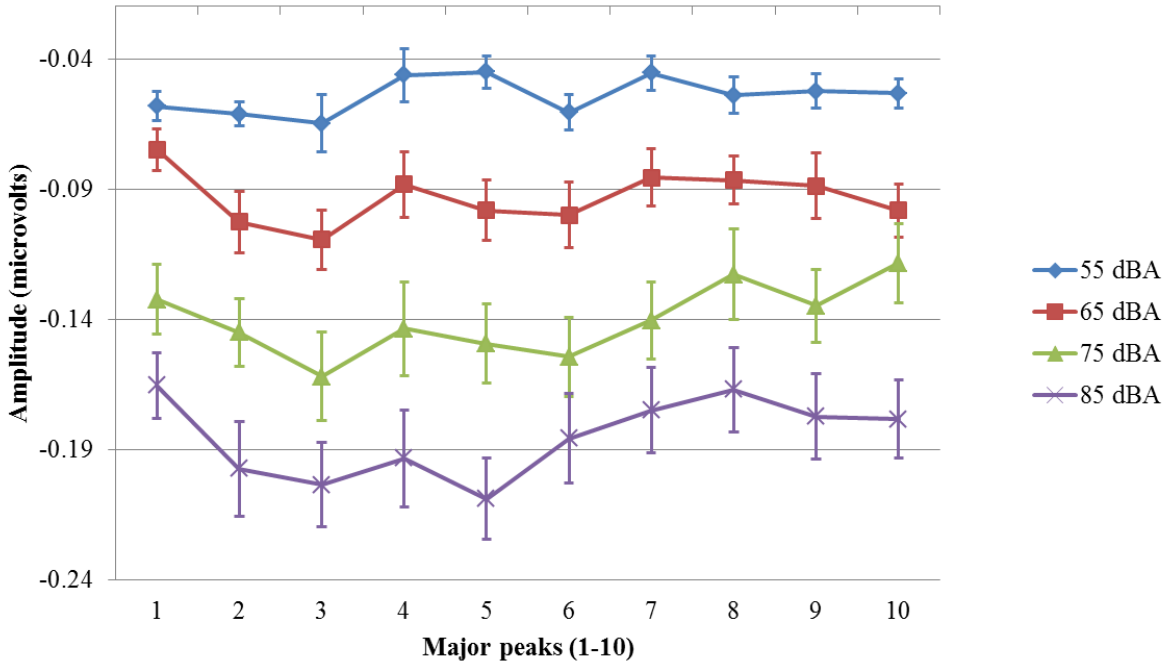

Figure 4.41: Major peak amplitudes as a function of sound level for the 100 ms /u/ vowel stimulus. [Study 2].

Table 4.27: Response latencies for all 100 ms vowel stimuli. [Study 2].

	Latency (ms) per sound level			
Stimulus	**55 dBA**	**65 dBA**	**75 dBA**	**85 dBA**
/a/ vowel	9.67	8.75	8.13	7.51
/ɔ/ vowel	10.00	8.96	8.33	7.61
/U/ vowel	9.38	8.75	8.13	7.40
/u/ vowel	10.83	9.69	8.85	8.02

latencies than males ($p<0.0005$) by an average of .80 ms.

The amplitudes corresponding to the major peaks in Fig. 4.40 are shown in Fig. 4.41. They are strikingly dissimilar to those of the other three stimuli. This is not surprising however as the frequency-domain analysis revealed monotonic increasing growth at F0 for the /u/ vowel in contrast to the other three stimuli, which appeared to show saturation and even decline at higher levels.

Lastly, the response latencies for all four 100 ms vowel stimuli are indicated in Table 4.27. Notice that they decrease with increasing level, as hypothesized. The latencies range from roughly 7-11 ms, which is consistent with the ABR and FFR latencies in the literature (see Section 2.3 and Appendix B). The vowels exhibit similar latencies at each respective level; however, the /u/ vowel in all cases exhibits slightly longer latencies than the other three stimuli.

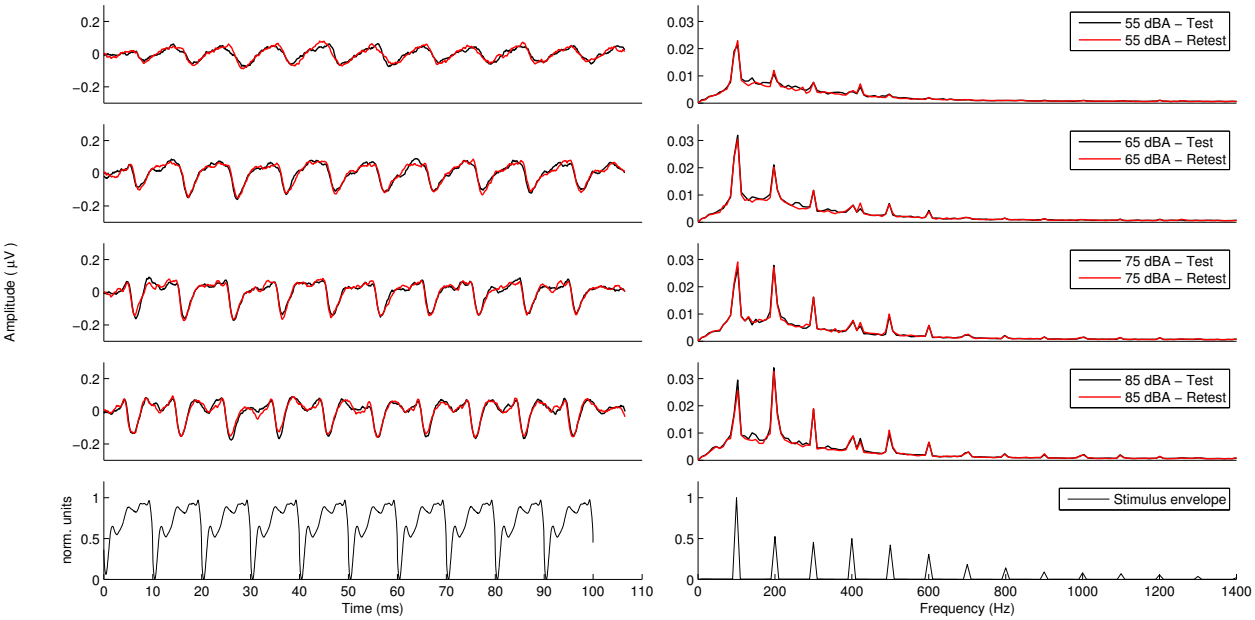

Figure 4.42: Test-retest reliability of the 100 ms /ɔ/ vowel eFFR. The test session grand-average (black) is nearly obscured by the retest session grand-average (red). [Study 2].

Characterizing the test-retest stability of the speech FFR

All four vowel stimuli yielded high test-retest reliability in both the time and frequency domain in the eFFR, as well as in the frequency domain of the sFFR. The waveforms of the sFFR do not exhibit nearly the same stability however. An example of this can be seen in the grand-average eFFR and sFFR plots of the test and retest conditions for the 100 ms /ɔ/ vowel in Figs. 4.42 and 4.43. The same plots for the remaining three vowels can be found in Appendix G, and all show a similar pattern.

Though the visual evidence is compelling, in order to quantify the strength of the test-retest reliability among the responses, the two correlative measures described in Section 4.2.5 were used for both the time and frequency domains, and in both the eFFR and sFFR. The results are tabulated in Tables 4.28 to 4.31. As an example here, Table 4.29 indicates that the Pearson's correlations between the amplitude spectra for the 100 ms /a/ vowel test and retest eFFRs averaged r=0.90 across all subjects and sound level conditions (highly stable!). Since the standard deviation of these intra-subject Pearson's correlations across all subjects and levels is 0.03, it appears as though this is very consistent phenomenon. This is further corroborated by the fact that the observed range of correlation values is .85-.95 here.

Although no statistical testing for differences was performed here, the results in Table 4.28 indicate that the correlations between amplitude spectra of the test and retest sessions are stable and roughly equivalent across level (although any differences seem to indicate a general increase in correlative strength with increasing level where they do appear). This

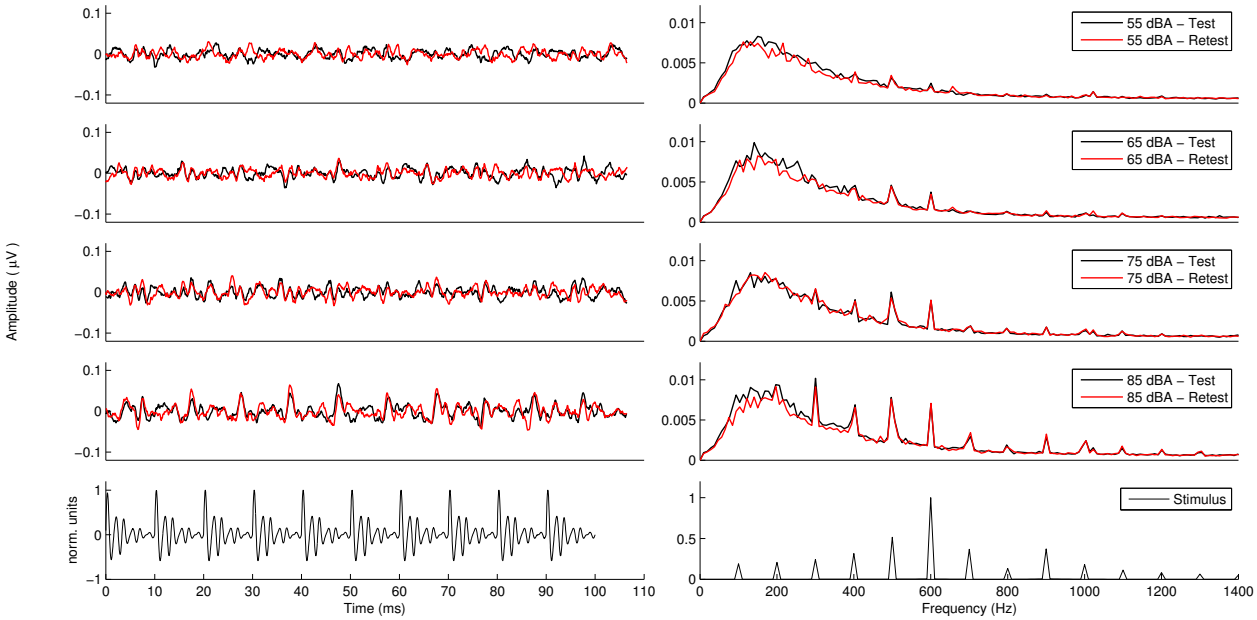

Figure 4.43: Test-retest reliability of the 100 ms /ɔ/ vowel sFFR. The test session grand-average (black) shows significant overlap with the retest session grand-average (red) in the area of the first and second formants in the amplitude spectra (right); however, the corresponding waveforms (left) appear to be markedly less correlated, which is confirmed in Table 4.30. [Study 2].

Table 4.28: Mean (SD) within-level intra-subject Pearson's correlations in the frequency domain (100 ms vowels). [Study 2].

	Vowel and response type							
Level (dBA)	**/a/ eFFR**	**/a/ sFFR**	**/ɔ/ eFFR**	**/ɔ/ sFFR**	**/U/ eFFR**	**/U/ sFFR**	**/u/ eFFR**	**/u/ sFFR**
55	.89 (.04)	.83 (.03)	.88 (.05)	.83 (.05)	.86 (.08)	.84 (.04)	.84 (.03)	.83 (.03)
65	.91 (.04)	.82 (.04)	.90 (.04)	.84 (.04)	.91 (.05)	.85 (.02)	.86 (.05)	.84 (.04)
75	.91 (.03)	.83 (.04)	.92 (.03)	.84 (.04)	.91 (.04)	.84 (.04)	.90 (.04)	.84 (.04)
85	.91 (.03)	.84 (.04)	.92 (.04)	.84 (.05)	.91 (.04)	.87 (.03)	.92 (.04)	.85 (.04)

Table 4.29: Mean (SD) intra-subject Pearson's correlations in the frequency domain, averaged across sound levels (100 ms vowels). [Study 2].

	Vowel and response type							
Measure	**/a/ eFFR**	**/a/ sFFR**	**/ɔ/ eFFR**	**/ɔ/ sFFR**	**/U/ eFFR**	**/U/ sFFR**	**/u/ eFFR**	**/u/ sFFR**
Mean	.90 (.03)	.83 (.02)	.90 (.03)	.84 (.02)	.90 (.03)	.85 (.02)	.88 (.03)	.84 (.02)
Range	.85-.95	.80-.86	.85-.96	.79-.88	.84-.96	.80-.89	.84-.94	.81-.89

Table 4.30: Mean (SD) within-level, intra-subject Pearson's correlations in the time domain (100 ms vowels). [Study 2].

	Vowel and response type							
Level (dBA)	**/a/ eFFR**	**/a/ sFFR**	**/ɔ/ eFFR**	**/ɔ/ sFFR**	**/U/ eFFR**	**/U/ sFFR**	**/u/ eFFR**	**/u/ sFFR**
55	.48 (.25)	.04 (.09)	.49 (.21)	.04 (.11)	.51 (.23)	-.01 (.10)	.21 (.18)	.02 (.10)
65	.63 (.20)	.05 (.14)	.61 (.17)	.06 (.13)	.63 (.20)	.09 (.17)	.42 (.26)	.09 (.17)
75	.66 (.19)	.09 (.14)	.70 (.17)	.17 (.16)	.67 (.18)	.17 (.22)	.56 (.24)	.02 (.22)
85	.67 (.17)	.20 (.20)	.68 (.20)	.31 (.22)	.63 (.24)	.26 (.24)	.70 (.16)	.17 (.24)

is in contrast to the results in Table 4.30, which seem to reveal significant increases with increasing level, and particularly so where the sFFR is concerned. This is consistent with what would be expected of general increases in signal-to-noise ratio with increasing level.

Table 4.31: Mean (SD) intra-subject Pearson's correlations in the time domain, averaged across sound levels (100 ms vowels). [Study 2].

	Vowel and response type							
Measure	**/a/ eFFR**	**/a/ sFFR**	**/ɔ/ eFFR**	**/ɔ/ sFFR**	**/U/ eFFR**	**/U/ sFFR**	**/u/ eFFR**	**/u/ sFFR**
Mean	.61 (.19)	.09 (.10)	.62 (.16)	.15 (.11)	.61 (.19)	.13 (.15)	.47 (.17)	.07 (.14)
Range	.27-.86	-.04-.45	.33-.86	-.02-.43	.18-.86	-.04-.60	.23-.77	-.05-.60

Tables 4.32 and 4.33 indicate the percentage of the time that a subject's test grand-average was maximally correlated with their own retest grand-average when correlations were performed with themselves and all other subjects. Chance performance here is 4.5% since there were 22 subjects in the study. Rather interestingly, the time domain representations result in higher average correlations than their frequency domain counterparts in all cases.

Table 4.32: Percentage of frequency domain test-retest correlations where the maximum correlation value occurs intra-subject (100 ms vowels). [Study 2].

Level (dBA)	**/a/ eFFR**	**/a/ sFFR**	**/ɔ/ eFFR**	**/ɔ/ sFFR**	**/U/ eFFR**	**/U/ sFFR**	**/u/ eFFR**	**/u/ sFFR**
55	31.82%	0.00%	22.73%	4.55%	31.82%	13.64%	4.55%	9.09%
65	50.00%	4.55%	31.82%	4.55%	36.36%	13.64%	18.18%	4.55%
75	36.36%	9.09%	31.82%	18.18%	40.91%	13.64%	45.45%	18.18%
85	40.91%	9.09%	45.45%	40.91%	50.00%	31.82%	45.45%	18.18%
Mean	**39.77%**	**5.68%**	**32.95%**	**17.05%**	**39.77%**	**18.18%**	**28.41%**	**12.50%**

Table 4.33: Percentage of time domain test-retest correlations where the maximum correlation value occurs intra-subject (100 ms vowels). [Study 2].

Level (dBA)	/a/ eFFR	/a/ sFFR	/ɔ/ eFFR	/ɔ/ sFFR	/U/ eFFR	/U/ sFFR	/u/ eFFR	/u/ sFFR
55	40.91%	4.55%	45.45%	9.09%	36.36%	4.55%	22.73%	4.55%
65	36.36%	9.09%	36.36%	13.64%	40.91%	18.18%	31.82%	9.09%
75	63.64%	13.64%	45.45%	22.73%	54.55%	31.82%	36.36%	9.09%
85	63.64%	22.73%	59.09%	40.91%	54.55%	31.82%	50.00%	27.27%
Mean	**51.14%**	**12.50%**	**46.59‘%**	**21.59%**	**46.59%**	**21.59%**	**35.23%**	**12.50%**

4.4 Discussion

The results of these two studies showed that effects of sound level can be observed via the speech-evoked FFR to vowel stimuli in the conversational range and that there are consistent patterns among these responses in the eFFRs and sFFRs in both the time and frequency domains. These effects are consistently observed regardless of duration, as the more abbreviated 100 ms vowel stimuli and the more natural 300 ms /a/ vowel stimulus displayed similar spectral growth and latency patterns with changing level throughout.

Frequency domain

The analysis of the amplitude spectra across all five vowel stimuli indicates that the auditory system exhibits population-wide phase-locking to frequency components in complex stimuli for both envelope cues via the encoding of F0 and its harmonics and temporal fine structure cues via the encoding of formant frequencies (specifically F1 and F2 here), consistent with temporal coding theories (see Appendix A). Indeed, both the eFFRs and sFFRs corresponding to each vowel provide solid estimates of the respective stimulus spectrums in all cases – which is well aligned with the notion that the FFR generally behaves as a good approximation to a lowpass filtered version of the stimulus itself, with the cutoff frequency located around the upper limits of phase-locking in the brainstem (~1500 Hz). Of course, owing to key response properties of the nonlinear auditory periphery, including synchrony capture, saturation, and phase locking to both TFS and ENV cues, the FFR should not generally be expected to have a perfect correspondence with such a lowpass version of the stimulus, and it bears mentioning that characteristics of such nonlinearities - such as enhanced F0 representation, for instance - are clearly present throughout.

The amplitude spectra of the sFFR measures of formant growth tended to increase with increasing level, consistent with our hypothesis and with the two previous studies by Krishnan (1999, 2002). Also consistent with these results was the more robust encoding of F1 harmonics in comparison to those of F2. Indeed, all statistical testing indicated significant

effects of level on F1, and the global measure of F1 resulted in significant pairwise comparisons between all means except in one single case. The analysis of the F1 centre harmonics revealed that this feature was not as robust by comparison to the global measure, however this should be taken cautiously from an applied perspective as the Holm-Bonferroni corrections are conservative and many of the p-values from the non-significant tests were smaller than p=.05. Regardless, the harmonics about F1 all tended to exhibit monotonic increasing growth, although frequent indications of amplitude saturation and a general nonlinear behaviour were observed. This latter effect indicated behaviour is consistent with the results of Krishnan (1999, 2002) and of many other AEP studies of sound level changes (see Section 2.3 and Appendices B and D).

Although the representation of F2 was clearly less robust than that of F1, significant effects of level were still consistently observed across all stimuli. The two 100 ms vowels with F2 centred at 900 Hz (/ɔ/ and /u/) in particular yielded significant pairwise comparisons between all sound level pairs via the global F2 measures (i.e. H8-H10), indicating consistent growth with level. H9 on its own consistently exhibited significant effects of level; however, it was relatively more prone to amplitude saturation effects, and fewer level-pair comparisons resulted in significant differences as a result. The F2 measures for both the 300 and 100 ms /a/ vowels were similar, with the H11-H13 measure exhibiting significant effects of level. All pairwise comparisons between levels for this measure were significant for the 100 ms vowel, and all but two were significant for the 300 ms vowel, which exhibited some amplitude saturation at the high (75 vs. 85 dBA) and low (55 vs. 65 dBA) ends.

Perhaps most novel among these results was the observation of the eFFR behaviour vis-à-vis the growth of the harmonic amplitude spectra with increasing level. Here, two primary observations were made – one concerning the nonlinear growth behaviour of F0, and one concerning the increasing harmonic richness via the representation of the higher harmonics present in the stimulus envelopes. With respect to F0, the first study clearly indicated that its amplitude growth pattern was nonlinear, as there was a mean increase from 55 to 65 dBA followed by a decrease to 75 dBA, and finally an increase to 85 dBA. While there was significant variation here between subjects here (itself a common occurrence (Coffey, Colagrosso, Lehmann, Schönwiesner, & Zatorre, 2016; Won et al., 2016)), this pattern was observed throughout aside from in the case of the 100 ms /u/ vowel, which exhibited a monotonic increasing growth profile.

This result is important since a good deal of the FFR literature deals with changes in F0 amplitude and so a clearer understanding of how F0 changes with level should help to provide context, and might enable a reanalysis of some results in this light, especially where comparisons are made between studies that used different presentation levels. It is also

particularly important here, as F0 amplitude was initially identified a key candidate feature for level prediction in ML and BCI-related tasks. These results clearly indicate that this is not generally viable.

The reason behind this observed F0 pattern is believed to be due to formant capture and the spread of synchrony (Krishnan, 1999, 2002; Yi et al., 2017; Young & Sachs, 1979). Since the healthy auditory system is able to capture the synchronous modulations related to the formant frequency closest to the characteristic frequency of auditory nerve fibers at low levels, spreading with increasing level primarily towards fibers with higher characteristic frequencies, fibers that are not representing formant frequencies in their temporal discharge patterns at low levels but rather are tending to be encoding modulations at the fundamental frequency of the stimulus are overtaken by the formant frequencies with increasing level, reducing the proportion of fibers representing F0 in their discharges (see Figs. A.16 and A.17 in Appendix A.7). Additionally, increasing level results in an increasingly large proportion of discharges to modulations at multiples of F0, resulting in relatively stronger representations of the harmonic envelope spectra above F0 and decreased temporal discharging to F0 itself.

As to why this type of behaviour was not observed in the responses to the 100 ms /u/ stimulus, two possible explanations which are not mutually exclusive can be offered. The most obvious from the perspective of the experimental setup is the fact that the /u/ vowel possesses the lowest F1 amongst the stimuli, with its F2 being centred at the lower of the two frequencies occurring in the stimulus set and being relatively small in amplitude. As a result, and consistent with the behaviour hypothesized above regarding F0 amplitude changes, it is possible that many higher characteristic frequency fibers within the phase-locking range and the range discernible via the FFR are increasingly able to discharge at F0 modulations (e.g. in the fibers between F1 and F2 not subject to formant capture and similarly in the fibers above F2) with increasing level.

The second possible suggestion is that the /u/ vowel clearly possessed the weakest response to the 55 dBA presentation level, and several subjects reported that this case in particular was relatively soft despite the A-weighting calibration that was intended to create similar loudness percepts between stimuli. Since the growth function in Fig. 4.41 appeared to indicate that the amplitude might be beginning to saturate, revealed by a decrease in the slope from 75 to 85 dBA, it might very well be that the behaviour observed in the four other stimuli would occur for this stimulus as well were higher presentation levels to be recorded.

The second of the primary observations referred to above pertains to the mentioned growth of the harmonic envelope spectra above F0. Increasing sound level generally resulted in increasing harmonic richness (i.e. growth of the combined RMS-amplitude of the harmonic envelope spectra above F0) for all five stimuli. Two exceptions occurred here – the most

significant being that the 100 ms /U/ vowel appeared to saturate from 65 through to 85 dBA from H2 to H6, with none of the associated pairwise comparisons attaining significance as a result. The other exception was the amplitude saturation of the 300 ms /a/ vowel from 75 to 85 dBA. Otherwise, all of the other pairwise comparisons were significant for all five vowel stimuli. This can in some sense be viewed as a proxy for the strong growth observed at H2, which yielded nearly identical statistical results with respect to the significance of follow-up tests, however H2-H6 appears to be the more robust measure and is aligned with qualitative observations of increasing harmonic richness with increasing level. As a result, the harmonics in this range or some RMS-amplitude combination of them are likely to be good candidate features for level prediction in an ML task.

Future studies to examine any possible correspondence between this harmonic richness and perceptual attributes such as the richness of sound would be interesting to pursue here in light of these results. For instance - might the enhanced representation of these higher harmonics of F0 be one of the reasons that music listeners turn up the volume to a desired level? Another related question is whether this increased harmonic richness contributes to intelligibility at higher sound levels. If so, then enhancing intelligibility - both for NH or HI populations in aided or unaided conditions - might require more than sufficiently increasing audibility.

In general, increasing level tended to result in FFRs that appear to provide increasingly good representations of the envelope and temporal fine structure cues of the stimuli that are known to be important for speech perception (namely F0, F1 and F2). This holds from a qualitative perspective as well, as simple visual inspections of the FFRs and their respective stimuli show increasing similarities with increasing level, with the caveat that effects of nonlinear transformations are clearly present. Some of these effects were detailed in the quantitative analysis performed above, and they are consistent with previous FFR studies of level encoding. Overall, estimates of effect size via η_p^2 revealed that anywhere from 50-90% of the variance in amplitude was due to level once sex effects were partialled out – which is substantial.

Although no attempts to observe or quantify suppression or distortion products were made, they are both likely to have impacted the data in addition to the mechanisms of formant capture and synchrony spread discussed above. Regardless, both vowel and level information are clearly represented in the FFR amplitude spectra of normal hearing adults, which is promising in that machine learning algorithms might be able to use such information for predictive purposes.

Time domain

The results of the time domain analyses were virtually identical across all five stimuli and are consistent with the literature. The response latencies all decreased with increasing level and fell within a range from of roughly 7-11 ms, consistent with the known conduction delays of the auditory brainstem. Response latency as a measure – despite its known clinical utility – should be interpreted rather cautiously here due to the fact that there are multiple contributing sources to the FFR spanning from the periphery to the cortex (Coffey, Herholz, et al., 2016; Tichko & Skoe, 2017) and frequency dependent effects are known to exist both in the actual latency encoding of stimulus information (e.g. higher frequency components activate the basilar membrane before lower frequency components do, all else being equal) as well as in the FFR amplitude itself as measured at a given electrode site due to constructive and destructive interference between the multiple contributing sources (Tichko & Skoe, 2017). As a result of this, it is not clear that a difference in response latency in the FFR corresponds to an actual difference in processing latency in the auditory system in general. Regardless, it is likely that the norms established here might be of clinical utility, since deviations from these ranges beyond the normal hearing population variability are likely indicative of pathology when the same stimulus, level, and recording setup are employed.

The analysis of the major peaks revealed that the eFFR robustly captures prominent peaks at the 10 ms pitch period of the stimuli (i.e. 1/100 Hz). The analysis of interpeak intervals indicates that tracking of periodicity is extremely precise and consistent across all presentation levels in these studies – a hallmark of the early auditory processing system. Since temporal processing is typically degraded in HI populations, extending this work to such populations in both aided and unaided conditions in order to better understand natural and processing-related changes in neural responses would be of interest. Additional studies along these lines via the study of effects of reverberation - particularly late reflections that smear the temporal waveforms (Al Osman et al., 2016, 2018) - would also be interesting to pursue.

Statistical testing on level-related latency differences among the peaks revealed significant effects of level for all stimuli, consistent with our hypothesis. Follow up tests further revealed significant differences between the latencies of each individual peak and between all pairwise comparisons of peak latencies between level conditions. These level differences are consistent with the results of Akhoun et al. (2008), which is the only previously published study on the effects of level on speech-evoked FFR latency. However, unlike Akhoun et al., we did not endeavour to analyze transient responses in the FFRs to any of the five stimuli, and so no comment on the relative latency changes between transient and sustained components can be made. This was done consciously due to the view that these transient responses were

not very reliable or robust during piloting, that they might be due to midbrain or cortical contributions such as the P20, and that they were not likely to contribute very meaningfully to the machine learning study (Sadeghian, 2012).

The amplitudes of the major peaks from both studies were also plotted for descriptive purposes and tended to reveal nonlinear and saturating effects, similar to the amplitude spectra of F0 which was verified to correspond (as expected) with the major peak latencies via bandpass filtering. The major peak amplitude of the 55 dBA presentation level was the smallest among all stimuli. The growth patterns here do not appear to be consistent with those observed in many other AEPs studied in the time domain (see Appendix B). This might be due to the differences in the stimuli typically employed, as the steady-state synthetic vowels here are significantly more complex than the clicks and tones typical of other such measures.

In support of this notion is the fact that although the two major level-related studies by Krishnan (1999, 2002) report increasing waveform amplitudes with increasing level (without providing actual objective measures), visual assessment of the reported waveforms seems to indicate a similar pattern to those observed here. In particular, the peak amplitudes between several of the 65 and 75 dB SPL conditions for various vowel stimuli appear very similar if not smaller at 75 dB (see Figs. D.5 and D.7 in Appendix D). Although Krishnan's observations are with respect to sFFRs as compared to the eFFRs analyzed here, a more recent result by Ananthakrishnan et al. (2016) shows very clearly that the eFFR of normal hearing subjects decreases in major peak amplitude between 75 and 80 dB SPL despite similar claims that "temporal waveform amplitude appears to increase with increase in stimulus intensity level". An identical claim related to the sFFR growth is made, which also does not appear to hold true between the 75 and 80 dB SPL cases. In any event, the amplitudes of the major peaks in the eFFR waveforms do not appear to be a very useful feature either for level prediction or for vowel prediction when the stimuli possess identical fundamental frequencies as they do here.

It should be mentioned that although the major peaks in Study 1 were selected to be the positive peaks, whereas negative peaks were selected in Study 2, the waveforms in Study 1 would have permitted the use of the negative peaks just as readily. That the positive peaks in Study 2 were nowhere nearly as easy to discern is likely due to the tripling in sampling rate (recall that the system used throughout had a fixed number of maximum samples regardless of stimulus duration), and so many smaller envelope features appear to be prominently represented in the 100 ms eFFRs whereas these same features are not visible in the eFFRs to the 100 ms /a/ vowel.

Unfortunately the sFFR waveforms did not permit any similar investigation of peak

latencies as consistent peaks tended to be obscured by the relatively high low-frequency noise floor. It may be possible to isolate formant contributions via bandpass filtering (using the Matlab 'filfilt' function to preserve absolute latencies, say) in order to enable within-vowel assessments of level-latency differences. This would be interesting to pursue further in future studies (and perhaps with a setup with less low-frequency noise) in order to determine whether or not the response latency changes differ between eFFRs and sFFRs and if there is a frequency-dependent interaction at play.

Sex differences in Study 2

Very few sex-level interactions were observed in the analysis of the amplitude spectra in Study 2. However, main effects of sex were observed in many of the amplitude spectra features, implying that general amplitude differences do exist between sexes. Although the mean spectra of females were in almost all cases equal to or greater than those of the males for the selected features under investigation, within-sex variation was substantial and so the confidence intervals for the respective male and female means almost always had some amount of overlap between them. Similarly, although the time-domain analysis revealed no sex-level interactions, main effects of sex on the normalized major peak latencies were observed for all four vowels, with the mean response latencies of the females always being shorter than those of the males (ranging from .64-.80 ms). In general, these results are very consistent with the literature (see Section 2.4) which seems to indicate that there are small effects of sex in various FFR measures across the population. Future studies interested in such effects should consider capturing head-volume data or some proxy thereof, as this is likely to account for some of the observed variation in latency and perhaps in the spectra as well (see Section 2.4).

FFR stability analysis

The analysis of the correlations both within and between subjects and both within and between sessions revealed that the FFR is generally an extremely stable measure. The eFFR and sFFR responses both possessed very strong amplitude spectra correlations within subjects, ranging from .79 to .96 across all four vowels in Study 2, for instance. The eFFR waveforms also had strong within-subjects correlations, although to a lesser extent, with values ranging from .18 to .86 in Study 2. The sFFR waveforms seemed to suffer here due to the low-frequency noise present, which generally resulted in either weak correlations or no apparent correlations at all. The inter-subject correlations in Study 1 were similar, although generally smaller than their intra-subject counterparts other than with respect to

the sFFR spectra, which is curious (although likely attributable to the removal of the phase information and the similar noise floors).

In general, the stability of the FFR within and between sessions indicates that the FFR is a good candidate for ML or BCI based applications that require consistent responses when faced with consistent stimulus presentations over time. This is particularly promising in light of the fact that the shorter 100 ms vowel durations selected for Study 2 are representative of vowel durations typical of conversational speech (Van Santen, 1992).

4.5 Conclusion

These two studies indicate that there are systematic changes in the representation of vowels in normal hearing adults via the FFR with increasing sound level. These changes manifest most evidently via the increased harmonic richness of the eFFR, via the increased representation of harmonics related to the vowel formants in the sFFR, and via stereotypical decreases in latency with increasing level in major peaks of the eFFR waveforms. These level-related changes in representation are stable both within and between recording sessions and are therefore likely to be of some utility both clinically and with respect to hearing related technologies such as hearing aids with a brain-computer interface component that helps to optimize for speech comprehension (e.g. vowel class) and appropriate loudness (e.g. level class).

Future work

Based on the observations of the studies reported above, future work recommendations related specifically to the study of the effects of level on the FFR include:

1. Repeating the above work with a full-EEG montage and a system capable of recording single trials.

2. Recording responses from ~60-95 dBA in finer level steps (3 dB steps, say) in order to gain a better sense of the level-related growth functions of the eFFR and sFFR in the time and frequency domains.

3. Utilizing more sophisticated and comprehensive testing of auditory abilities, such as tests of musicality, listening mode, hidden-hearing loss, speech-in-noise testing, etc.

4. Acquire loudness assessments from all subjects for all of the stimulus-level conditions in order to allow for level-loudness analysis.

5. Perform a similar study on hearing impaired populations in both aided and non-aided conditions, and include assessments of sensation level and loudness. Extensions related to the effects of reverberation should also be considered here.

6. Consider including assessments of context and transients using consonant-vowel stimuli (e.g. /da/, /ga/, /ba/, /du/, /gu/, /bu/, etc.). The representation of natural vowels with varying pitch and pitch tracks typical of males, females, and children would also be of interest here.

Chapter 5
On the automatic classification of the FFR using machine learning

5.1 Introduction

This study addresses the following research question:

> *Can machine learning methods discriminate between FFR responses to different vowels AND/OR levels with good accuracy?*

This is an important question that has implications with respect to our understanding of what exactly is being indexed in this ensemble electrophysiological response, as well as to our notions of how the FFR might be exploited in clinical and technological applications.

To our knowledge, only three studies designed to assess FFR classification performance from a statistical modeling or ML perspective have been published to date. All three of these have sought to assess vowel classification, and have differed significantly in design. The most similar approach to the one taken here is that of Sadeghian et al. (2011, 2012, 2015), wherein responses to five 300 ms English vowels at a single sound level of 80.5 dB SPL were recorded from 8 normal hearing adults, and the resulting sustained spectral features (i.e. eFFR and sFFR harmonics of F0) and transient features were used to train linear discriminant analysis (LDA) and support vector machine (SVM) classifiers. Classification accuracies as high as 83.3% were obtained (with chance performance being 20%), and it was determined that the harmonic features were by far the most crucial contributors to model performance. In their study, each FFR was averaged over 500 trials and six responses were recorded per vowel per subject. All data was captured during a single session.

In contrast, a study by Yi et al. (2017) examined the ability to decode single-trial FFRs collected from 38 normal hearing subjects (30 females) to two 250 ms vowels each produced by two speakers. The responses were then projected onto a low-dimensional feature space derived from the same two vowels being produced by 40 separate speakers. A supervised ML classifier, Gradient Boosted Decision Tree, was then trained on a subset of FFRs from each subject, and tested on the remaining responses. Vowel classification performance yielded a mean of 66.8%, and differed significantly from the chance performance of 50%. The spectral features used by the classification model closely corresponded to the formant structure of the

stimuli. Although the performance is modest and the classification task was binary, the fact that a single trial possesses sufficient, interpretable information for such discrimination has significant implications with respect to future technological development, as the time cost of averaging can be substantial and could foreseeably limit the uptake of any technology that relies upon a cumbersome, repetitive recording process.

Finally, a study by Won et al. (2016) obtained FFRs using four different 70 ms vowels presented at 80 dB SPL in 38 (28 female) normal hearing listeners in addition to vowel recognition data in order to compare behavioural and neurophysiological responses. Their 'multidimensional phoneme identification' model (due to Svirsky (2000)) relied on input features derived from frequency difference measures between the formants present in the acoustic stimulus itself and those present in the FFR in order to predict vowel confusion patterns. These were then compared with perceptual confusion patterns. Their data revealed that there were large variations in formant encoding as indexed by the FFR, and that these variations were related to recognition performance in the behavioural task, and so they were able to successfully predict good (>50% vowel recognition) versus poor (<50%) vowel identification performers with ~100% accuracy based on the FFR alone.

Two additional FFR studies warrant mention here: the first is due to Jafarpisheh et al. (2016), and the second to Bidelman et al. (2013). Despite the fact that it only sought to determine the utility of a nonlinear feature extraction technique for the oft-used consonant-vowel stimuli /ba/, /da/, and /ga/ and did not endeavour to actually test classification performance, Jafarpisheh et al. (2016) showed that there are clearly important gains to be made in the pre-processing and feature extraction components of the ML pipeline with respect to FFR classification. Using data captured from normal hearing adults, a recurrence method from the field of nonlinear dynamics that looks at trajectories in phase space was employed. The feature extracted essentially indexes for state changes in the formant-transition period, which is known to contain a great deal of differential information among these stimuli (K. L. Johnson et al., 2008). The feature extraction process as performed on the collected FFR data results in highly separable classes, despite the fair amount of variance in their sample, which is common to FFR data. This is significant not only in that it establishes the possibility that advanced feature extraction processes might enable the accurate classification of consonant-vowel stimuli as represented in the FFR – as consonant discrimination is often particularly problematic for hearing impaired persons (Phatak, Yoon, Gooler, & Allen, 2009) but also by virtue of the fact that the process itself is known to perform well in noisy environments.

The study by Bidelman et al. (2013) – described in Section 3.4 above – modeled the cortical transform of the first formant (F1) frequency as detected by ERPs, and showed that

their model possessed a good deal of agreement with the actual recorded cortical ERPs. Their study elicited FFRs that differed in F1 location along a continuum between /u/ and /a/ vowel classes, and their modeled cortical ERPs were good predictors of a subject's categorical speech perception. Since the transform that fit their data was sigmoidal in nature, it is basically analogous to an ML implementation of a one-vs-one, single-feature logistic regression model (see Section 5.2.1) aside from the fact that it is effectively 'trained and tested' on the entire dataset.

Taken as a whole, these studies support our hypothesis that supervised learning models trained on the dataset obtained via Study 2 (see Chapter 4.3.2 for a full description) will permit the accurate prediction of vowel class. Furthermore, as the analysis in Study 2 demonstrated, there are statistically observable level effects among the responses, both in the time and frequency domains, and so level class predictions are expected to perform significantly above chance as well – something that has not yet been shown to our knowledge.

Based on said analysis, it is expected that eFFR features will be of greatest importance for level classification tasks both in the time domain as systematic shifts in peak latency were observed (and there are qualitatively observable waveform differences as well), and in the frequency domain where there was an observable increase in the harmonic growth or richness of the eFFR with increasing level. These same spectral envelope features are not expected to greatly facilitate the differentiation between vowel classes due to the extreme similarity between the response (and stimulus) envelope spectra.

The sFFR features are expected to contribute significantly to vowel classification performance, since the F1 locations differ from vowel to vowel, and there are two different F2 peak locations among the four vowels. These same features tended to exhibit growth with increasing level, so the sFFR is also expected to contribute meaningfully to level class prediction performance. Since the sFFR waveforms have a relatively high noise-floor in the low frequency range, it is not expected that they will contribute greatly to classification efforts in their raw form; however, pre-processing of waveforms via filtering is expected to enhance both vowel and level class differences.

The analysis in Study 2 also suggests that using sex as a feature might improve classification results, and so its utility is evaluated as well.

Lastly, since the effects of level observed in Study 2 were not always monotonic in nature, we propose that vowel classification should not be too dissociated from level classification, especially when being considered in an applied context. This has not been addressed in any of the previous studies mentioned.

By testing these hypotheses using a variety of well-known ML classifiers, as well as two novel 'signal-similarity' algorithms designed to exploit the high intra-subject and within-class

correlations characterized in the analysis of Study 2, a variety of classification benchmarks will be established. The ability to accurately predict speech and/or level class as indexed via normal FFRs in response to stimuli typical of conversational speech such as the 100 ms vowels used here, and to further predict behavioural performance – as available converging evidence suggests might be possible – would certainly reveal the potential for a host of novel clinical and technological applications. In addition, a better understanding of the various features and their relative importance in these predictive tasks may result in a deeper understanding of healthy and pathological auditory function.

5.2 Methods

The methods used to capture the data utilized for this study are described in detail in Section 4.2 above. The raw (or unfiltered) waveforms and spectra that were processed in Study 2 using Matlab R2014a (v.8.3.0.532, by MathWorks, Natick, MA) were all imported into Python (v. 3.5.2) in order to make use of the open-source scikit-learn (v. 0.18.1) machine learning library. In some instances – as indicated – the time series data was filtered using fifth order Butterworth bandpass filters with cutoffs of 65 and 750 Hz for the eFFR, and 365 and 1350 Hz for the sFFR. Additionally, the time series data was normalized by detrending and dividing by the standard deviation in some cases, as indicated. Finally, some algorithms were tested using each subjects combined test and retest subaverages for each vowel-level condition in order to determine if the boost in signal-to-noise ratio results in an increase classification accuracy.

5.2.1 Classifier methods

Several well-known classifiers were used, both in order to determine what the best candidate models are for any future implementations of such tasks (e.g. in a BCI), and also to gain insight into why their performance differs in instances where it does to the extent that this is possible. For instance, does a linear discriminant analysis (LDA) classifier work best for level prediction, and if so, does this performance depend relatively more on envelope or temporal fine structure features? Do classifiers trained on time-series data outperform those trained on spectral data?

The classifiers that were utilized in this study, along with a brief description, are presented below (see *The Elements of Statistical Learning* by Hastie, Tibshirani, and Friedman (2009) for a good introduction to each model). Each of the conventional classifiers were implemented using scikit-learn in their default states unless otherwise indicated.

Linear discriminant analysis (LDA): A classifier with a linear decision boundary, separating classes from one another by creating hyperplanes where the ratio of between-class variance to within-class variance is maximized, thereby optimizing class separation. LDA assumes that the features are normally distributed, and it reduces the dimensionality from the original number of features to C-1 features, where C represents the number of classes.

Support vector machine (SVM): Similarly to LDA, SVM creates hyperplanes to separate classes from one another, and the data points closest to the hyperplanes are referred to as support vectors. Whereas LDA finds the hyperplane that best separates all of the data points, SVM finds hyperplanes that best separate only the points along the frontier between classes. Unlike LDA, SVM can work well even if the data is not linearly separable in the base feature space, and it performs well in very high-dimensional spaces. Various linear and nonlinear kernels (i.e. similarity measures) are available, which are often selected depending on the size of the dataset, computing costs, and the ratio of the number of features to the number of observations.

Logistic regression (LR): A well-known regression model that is utilized when the dependent variable is categorical (typically binary), and performs well even if the features are correlated. Here, the feature space is mapped to a number between 0 and 1, with a typical threshold being placed midway at 0.5; however this threshold can be adjusted to optimize for various outcomes (e.g. minimizing false negatives for lethal diseases). Multiclass logistic regression implements a one-vs-all classification scheme, and therefore trains C models for C classes. Unlike models such as SVM or decision trees, logistic regression allows for a clear probabilistic interpretation, indicating that the presence of a feature (or risk factor) increases the odds of a given outcome by some specified factor.

K-nearest neighbours (KNN): A classifier that simply predicts class membership based on the most common class among the k-nearest neighbouring points in the feature space. For k=1, the test data point is simply predicted to belong to the class of the single nearest neighbouring training data point. Changing the k-value can radically change the decision boundary. It can produce strong results when there are multiple clusters in the feature space all bearing the same class label, and is robust to noisy data. However it is computationally expensive, as it requires the computation of distance for each test sample to all of the training samples.

Decision tree (DT): The decision tree is a rule-based classifier that creates a tree graph to go from observations about a test data point, represented via the tree branches, to conclusions about its likely label, represented by a leaf node. Decision trees are simple to understand and interpret, but they can create overly complex classifiers that do not generalize well (i.e. they are prone to overfitting). Decision trees can be unstable, as completely different trees

can often be generated depending on the sampling of the dataset.

Random forest (RF): The random forest classifier is an ensemble ML method that creates a 'forest' full of decision tree models created on various subsets of the dataset. The classifier predicts that the test data belongs to the class label that has the highest mean predicted class probabilities over the ensemble of decision trees. They are powerful tools since they tend to mitigate overfitting issues related to decision tree models and improve classification accuracy.

The two additional signal-similarity algorithms that were created in order to exploit the high test-retest correlations within subjects, as well as between subjects within vowel and level classes (see Section 4.3.2), are described below.

Signal subtraction algorithm:

1. For each 'test' condition signal, subtract the signal from all available signals in the 'retest' set, and take the root-mean-square of each difference, thereby producing a single value associated with each 'retest' signal (with more similar signals resulting in smaller values).

2. Take the most frequently occurring label among the n-smallest values produced, and predict that this is the correct label for the 'test' signal.

3. Select the n value that yields the best predictive accuracy (NB: this may vary depending on predictive class of interest – i.e. whether or not vowel, level, or vowel-level class predictions are sought).

Here, the 'test' condition signals refer to the eFFR or sFFR time series, but it is possible to test their amplitude spectra, or a signal formed by concatenating any combination thereof.

Signal correlation algorithm: A nearly identical algorithm to the above, the correlation algorithm simply correlates each 'test' signal with every 'retest' signal, and then selects class labels based on the most frequently occurring among the n-largest correlation values. Here again, the n-value that yields the best predictive accuracy is selected, and this may vary depending on the predictive class of interest.

5.2.2 Model training, testing, and evaluation methods

The ML work performed in this study can be broken down into three sections:

1. the application of common classifiers on the harmonic spectral features of the FFRs

2. the application of an SVM to the time series data of the FFRs

3. the application of the signal-similarity algorithms to the time series data of the FFRs, split into their 'test' and 'retest' conditions

Throughout, models were trained and tested on level class, vowel class, and level-vowel class labels, as indicated. The first section tested the first thirteen harmonic features (i.e. F0 to H13) separately for eFFR and sFFR measures (13 features) and combined over both eFFR and sFFR measures (26 features). This was typically performed using the entire dataset, and at times by utilizing the subjects' averaged FFRs from their test and retest conditions across all vowel-level classes (i.e. both of their 55 dBA /a/ vowel responses from the 'test' condition were averaged, and so on). The former is referred to throughout as 'all-subaverages', whereas the latter is referred to as 'combined-subaverages' for clarity.

For training and testing purposes, a 10-fold cross-validation (CV) was employed as it partitions all of the data into 10 different sets (with equal class representation in each), training 10 separate models, and tests each data point exactly once in order to provide a balanced assessment of overall performance (Hastie et al., 2009, Ch. 7). In addition to the 10-fold CV, a leave-one-out (LOO) CV paradigm was also tested for sake of comparison – both with the 10-fold CV results as well as to the results of Sadeghian et al. (2015) as they employed this method in their study. The LOO is generally only utilized on very small datasets as it is computationally expensive (a new model needs to be trained for every single FFR), and the models are prone to increased complexity or overfitting. As a result, such models are not typically representative of the type of performance a production model (i.e. one deployed to maximize general performance in a functional product) would possess (Rao, Fung, & Rosales, 2008).

A leave-one-subject out (LOSO) paradigm was also assessed, wherein the data from one subject was held out for testing purposes, and the data from all of the remaining subjects was used to train a model. This was done in order to determine how a system that had no prior knowledge of an individual's responses might perform. Two additional feature sets were tested in this section: one in which a categorical 'sex' feature was added to the feature set in order to determine if resulted in any meaningful changes in performance, and one in which amplitude spectra of the two points on either side of each harmonic were included as features in order to determine if the additional spectral information (i.e. noise) impacted performance.

Wherever KNN classifier results are presented based on the 10-fold CV method, the k-value that produced the results will be indicated. Since the optimal k for each subject or for each separate test data point varied, no k value is reported for the LOO or LOSO paradigms; however k was restricted at all times to a range from 1-40 in these instances.

All instances of the classifiers in this section were default implementations, with the

LinearSVC kernel being used for all SVMs. This kernel uses a one-vs-the rest approach to multiclass classification, and so C models are trained for C classes.

In the second section, SVMs using a nonlinear radial basis function kernel (implemented using scikit-learn's NuSVC class) were trained and tested on the time series data, both in its raw and filtered forms, and on both the all-subaverages and combined-subaverages datasets. Again, the eFFR and sFFR waveforms were used both individually and in combination. Here, a one-vs-one approach is employed, and so C(C-1)/2 models are trained for n classes. Since SVMs are known to perform well when the ratio of feature sets to observations is relatively large, this model was selected after some piloting as the best candidate.

Finally, in the third section, the signal-similarity algorithms were run by comparing each test response to each retest response both on an intra-subjects basis wherein the test and retest sets represent responses from each subject alone, and the performance indicated is computed as the mean over all subjects, and across all subjects wherein each 'test' response is compared to every 'retest' response available.

Throughout, classification accuracy is computed as the aggregate of correctly classified test samples over the total number of test samples. Since there are four vowel classes and four sound level classes, these respective categorization tasks possess a chance accuracy level of 25%; for vowel-level tasks (e.g. predicting '55dBA /a/ vowel'), the 16 unique classes result in a chance accuracy of 6.25%. In addition to classification accuracy scores, a confusion matrix showing a cross-tabulation of the actual test classes versus the predicted class instances is often computed in order to give a more subtle sense of performance since accuracy scores can seem to be excessively punitive if there are a large number of classes, and it is often important to understand the type and magnitude of classification errors being made.

5.3 Results

5.3.1 Using the FFR amplitude spectra of F0 and its harmonics as features (F0 to H13)

The relative vowel-level classification performance of the classifiers trained on all of the eFFR and sFFR harmonic features using 10-fold CV is displayed in Fig. 5.1. The LDA model is the best performer, with a classification accuracy of 32.5%, which is 5.2-fold increase over the chance performance.

The heat map for the confusion matrix corresponding to the LDA model's output, shown in Fig. 5.2, reveals that the majority of predictions occur along or near to the diagonal, and so a large number of the misses are 'near-misses'. A closer inspection indicates that a good

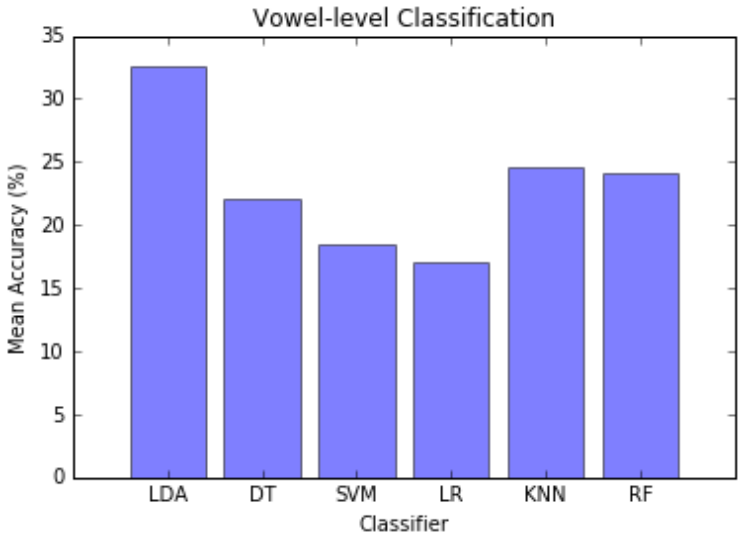

Figure 5.1: The vowel-level class predictive performance of each classifier trained and tested on both the eFFR and sFFR sets of harmonic features is shown in terms of the mean accuracy scores obtained using 10-fold CV.

proportion of these (50.7%) are due to predictions that are within the appropriate vowel class, but are off in level-class by ±10dB. Also clearly visible is the fact that many of the 55 dBA test cases result in a high number of vowel class confusions, and that certain classes seem to be much easier to classify (e.g. the 55 dBA /U/ class), whereas others prove to be more difficult (e.g. the 65 dBA /ɔ/ class). Notice too that there are many vowel-class confusions that nonetheless properly predict sound level. This is quite visible among the 85dBA /u/ class predictions, for instance.

The results of various input feature combinations across all classifiers for this vowel-level task are summarized in Table 5.1. Regardless of the features used here, LDA outperforms all other models. The addition of sex as a feature resulted in a classification accuracy of 32.4%, which is 0.1% worse than when the feature was not included, indicating that it does not meaningfully impact performance.

The relative performance of each classifier on the vowel categorization task trained and tested on the sFFR spectral features alone via 10-fold CV is displayed in Fig. 5.3. Although both the KNN and DT models performed nearly as well, the random forest classifier achieved the best accuracy by correctly classifying 71.7% of the vowels regardless of their level class. This is in line with the trend seen above in Fig. 5.2 wherein a large proportion of the classification errors occurred within vowel classes.

The confusion matrix (Fig. 5.4) corresponding to the random forest classifier's results indicate that the misclassification rates are fairly balanced between vowel classes; however there is a clear bias towards confusions occurring with the /a/ vowel class in comparison to the others.

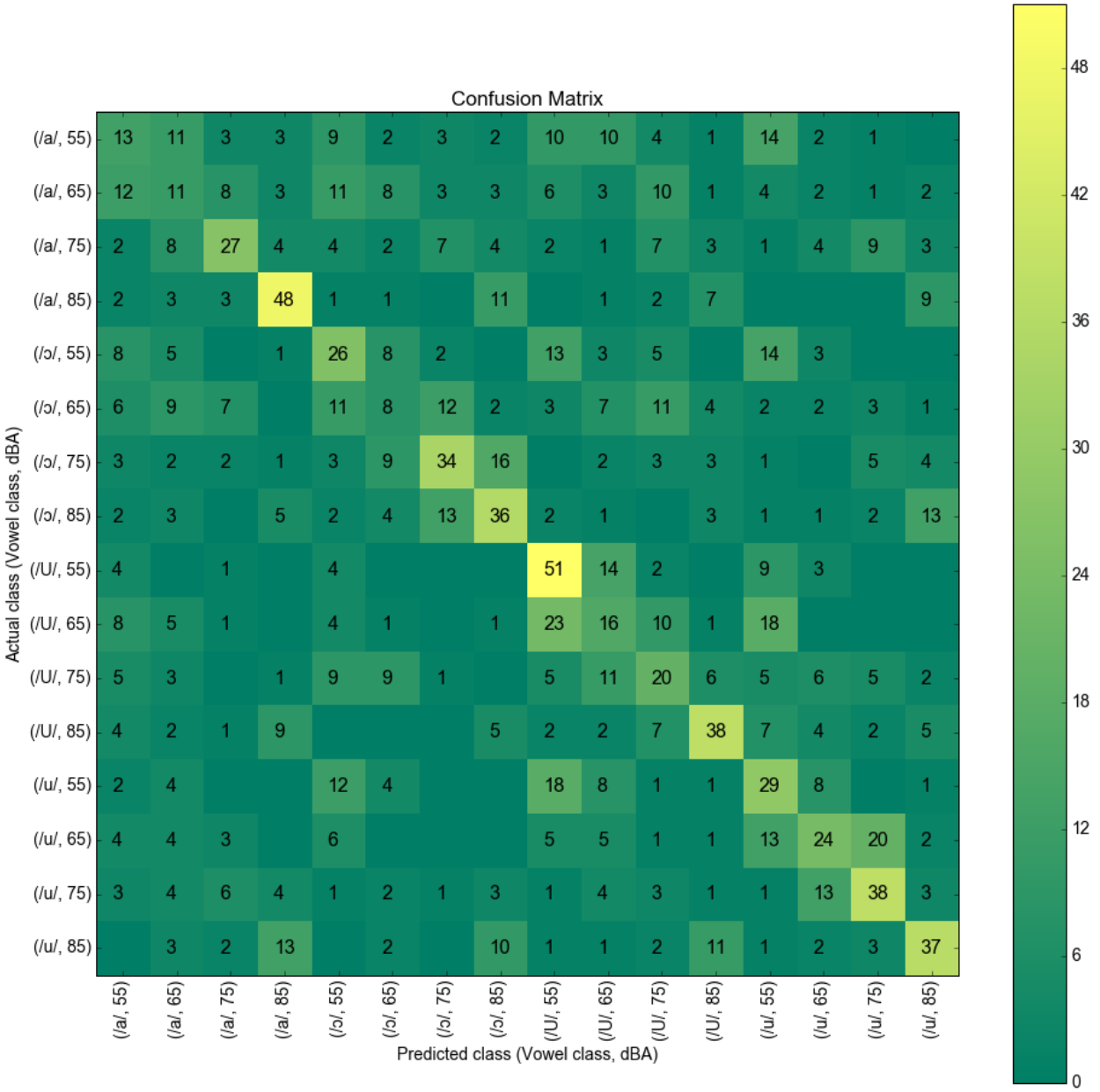

Figure 5.2: The confusion matrix corresponding to the LDA classifiers performance on the vowel-level classification task using both the eFFR & sFFR sets of features, with the actual vowel and level class pairs indicated on the ordinate and the predicted class pairs on the abscissa.

Table 5.1: The results of the classification performance on the vowel-level task obtained using F0 and its harmonic features (F0 to H13) and the 10-fold CV method. The feature sets used include the use of the eFFR and sFFR harmonics on their own and in combination. The best performance in each instance is indicated in bold.

		Classification accuracy (%) per feature set		
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	6.25	**32.5**	**25.2**	**15.7**
SVM	6.25	18.3	15.9	12.4
LR	6.25	17.1	15.6	12.2
KNN	6.25	24.4 (k=39)	20.0 (k=33)	11.6 (k=1)
DT	6.25	22.4	17.4	13.3
RF	6.25	24.7	22.0	13.7

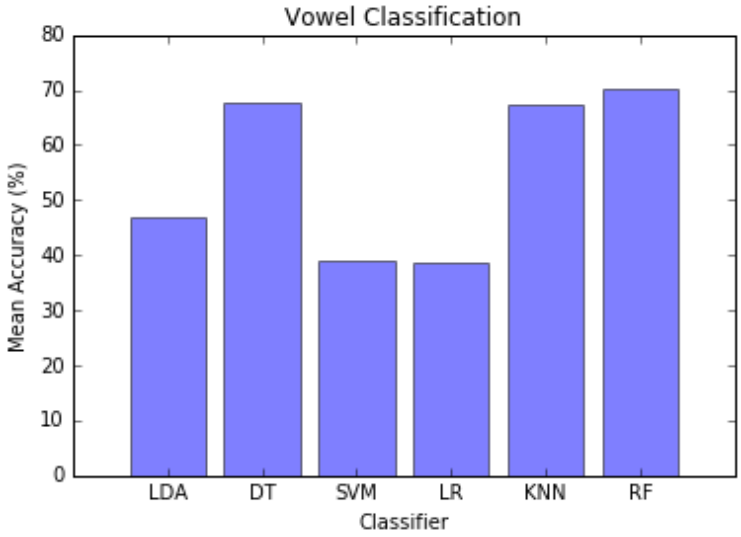

Figure 5.3: The vowel class predictive performance of each classifier trained and tested on the sFFR-only harmonics feature set is shown in terms of the mean accuracy scores obtained using 10-fold CV.

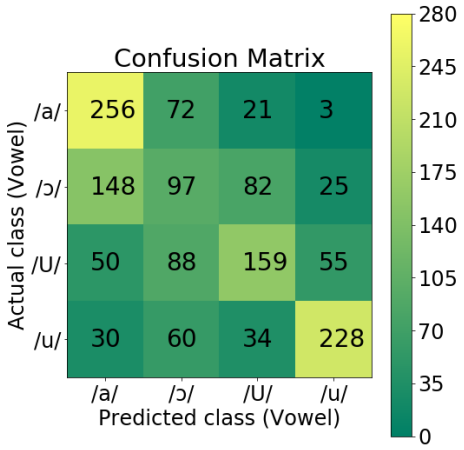

Figure 5.4: The confusion matrix corresponding to the RF classifier's performance on the vowel categorization task, trained and tested using 10-fold CV on the sFFR-only harmonics feature set.

Table 5.2: Classification accuracies for the vowel-class prediction task obtained using the 10-fold CV method.

Algo.	Chance (null set)	Classification accuracy (%) per feature set: eFFR & sFFR	eFFR only	sFFR only
LDA	25	**54.8**	**43.2**	47.0
SVM	25	38.8	37.0	38.9
LR	25	38.9	36.9	38.8
KNN	25	42.5 (k=18)	38.5 (k=30)	67.4 (k=1)
DT	25	43.6	34.6	67.5
RF	25	48.6	40.1	**71.7**

The results of various input feature combinations across all classifiers for this vowel prediction task are summarized in Table 5.2. Although the LDA classifier outperforms the others when all eFFR and sFFR harmonics are used as inputs, as well as when only the eFFR features are used, the RF classifier clearly obtains the highest accuracy score when using the sFFR harmonics alone. Both the DT and KNN classifiers achieve good performance here as well at ~67.5% each. The addition of sex as a feature resulted in a classification accuracy of 70.6% when using the RF classifier with only sFFR harmonics as features, which is roughly 1% worse than when the sex feature is omitted.

The relative performance of each classifier on the level categorization task trained and tested on both the eFFR and sFFR harmonic features using 10-fold CV is displayed in Fig. 5.5. Although both the KNN and RF models performed nearly as well, the LDA classifier achieved the best accuracy by correctly classifying 52.6% of the level classes regardless of their vowel class. This is in line with the trend seen in Fig. 5.2 wherein a large proportion of predictions correctly identified both vowel and level classes, but many of the vowel-class confusions correctly predicted level class membership (particularly at 55 and 65 dBA). Using sex as a feature resulted in a 51.7% classification accuracy when tested on the LDA classifier with all harmonic features being utilized, which once again represents a slight reduction in performance of roughly 1%.

Confusion matrices for each of the LDA and KNN classifiers are presented in Fig. 5.6 (left and right, respectively). Although they yielded nearly identical classification accuracies, there are some noticeable differences in the types of confusions made, which might warrant consideration if either algorithm was to be implemented in an operational system. For instance, the LDA classifier very rarely confuses a 55 dBA response with an 85 dBA response, whereas the KNN classifier produces ~4.7 times as many errors of this type. The LDA classifier generally exhibits better performance at the 75 and 85 dBA levels, which could be particularly important in a system that seeks to minimize any excessive experience of

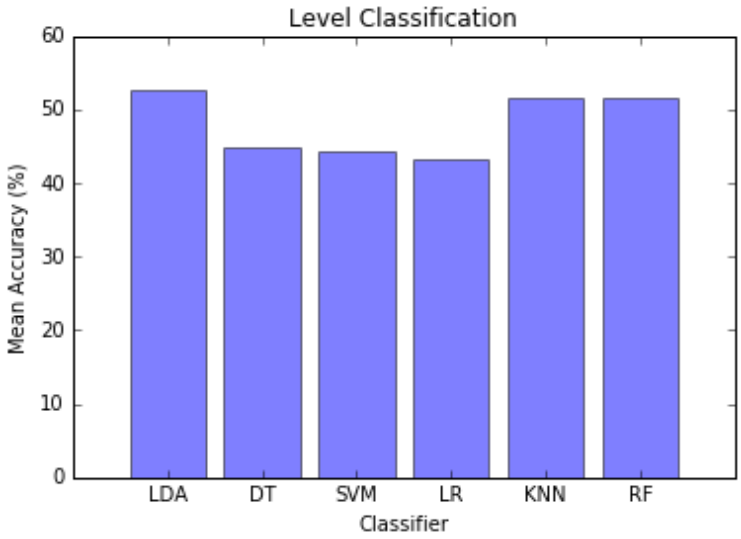

Figure 5.5: The level class predictive performance of each classifier trained and tested on the eFFR and sFFR harmonics feature set is shown in terms of the mean accuracy scores obtained using 10-fold CV.

Table 5.3: Classification accuracies for the level-class prediction task obtained using the 10-fold CV method.

	Classification accuracy (%) per feature set			
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	25	**52.6**	**47.9**	**35.0**
SVM	25	44.2	41.1	34.2
LR	25	43.2	41.3	34.0
KNN	25	51.6 (k=22)	46.5 (k=37)	33.0 (k=39)
DT	25	42.6	41.0	28.3
RF	25	51.3	47.7	31.0

loudness, for instance. The overwhelming majority of predictions lie within 10 dB of the actual level for both the LDA and KNN classifiers (86.6% and 84.6%, respectively).

The results of various input feature combinations across all classifiers for this level prediction task are summarized in Table 5.3. Although the LDA classifier outperforms the others in general, the RF and KNN classifiers both yield very similar results. The addition of sex as a feature here resulted in a classification accuracy of 51.7%, which once again represents a ~1% reduction in classification accuracy in comparison to the model that omits the feature.

The results for all of the same classification tasks performed above using a LOSO CV approach rather than the 10-fold CV approach used above are summarized in Tables 5.4 to 5.6. In addition to the mean classification accuracies, the range of subject performance is also indicated. The general performance here tends to be comparable; however there is a substantial decrease on the maximum classification accuracy of the vowel-class prediction task. Whereas the random forest classifier attained a classification accuracy of 71.7% in

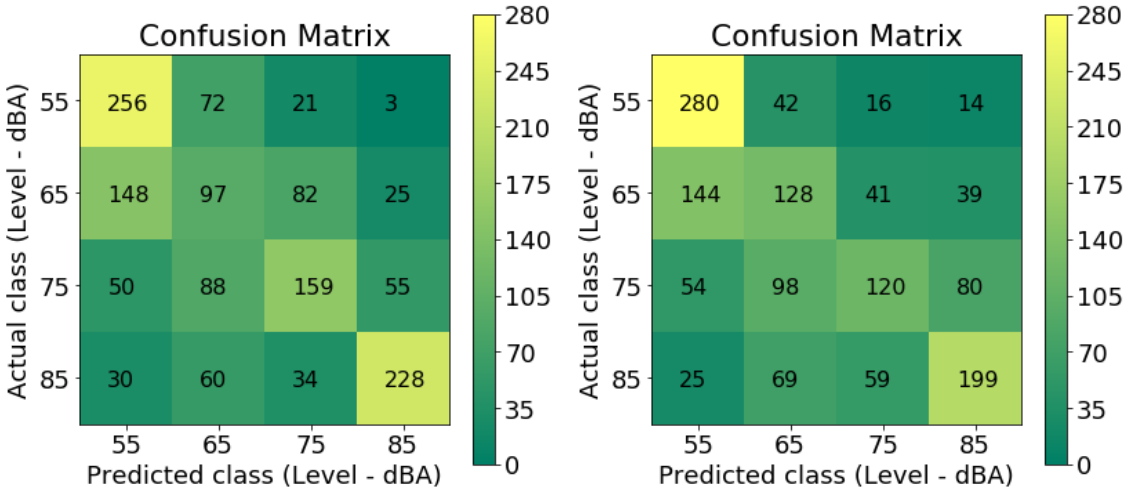

Figure 5.6: Confusion matrices corresponding to the level categorization performance of the LDA (left) and KNN (right) classifiers trained and tested using 10-fold CV on the eFFR & sFFR harmonics feature set.

the 10-fold CV paradigm (see Table 5.2), a maximum accuracy score of 53.6% was observed using the LOSO approach (see Table 5.6). It is worth noting that the performance of the DT and RF classifiers dropped off substantially in the vowel classification task when the features were restricted to the sFFR harmonics. This implies that the rules created by the models trained on all of the other subjects did not generalize well to 'new' subjects.

Additional results acquired by using the LOO approach, as well as from the 10-fold CV approach on the combined-subaverages are detailed in Appendix H. Combining the subaverages did not result in any marked increase in predictive accuracy here. The LOO results were also comparable to those described above; however the maximum classification on the vowel task was 55.6%, which is similar to the LOSO result of 53.6% (see Table 5.5) and roughly 17% lower than the maximum accuracy obtained using 10-fold CV (see Table 5.2).

5.3.2 Using the FFR time series data as features

The results of the SVM classifiers trained and tested on the filtered and normalized eFFR and sFFR combined-subaverages time series data (used throughout this section) are summarized in Table 5.7. The greatest classification accuracies attained for vowel-level class prediction was 50.9%. In this instance, both the filtered and normalized eFFR and sFFR waveforms were used as the feature set. A maximal vowel-class prediction accuracy of 74.6% was attained using the sFFR waveform alone. Finally, a level-class prediction accuracy of 61.6% was attained using both the eFFRs and sFFRs as features.

The results obtained by the addition of sex as a feature are shown in Table 5.8. The

Table 5.4: Mean classification accuracies and the range of performance among subjects on the vowel-level prediction task using the LOSO CV method.

		Classification accuracy (%) per feature set		
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	6.25	**29.7 (14.1-50.0)**	**23.6 (9.4-48.4)**	**15.6 (6.3-26.6)**
SVM	6.25	17.2 (9.4-26.6)	15.9 (9.4-26.6)	12.4 (4.7-20.3)
LR	6.25	16.3 (7.8-25.0)	15.3 (7.8-25.0)	12.5 (4.7-20.3)
KNN	6.25	25.6 (10.9-42.2)	22.8 (9.4-39.1)	14.3 (6.3-23.4)
DT	6.25	17.5 (7.8-32.8)	14.2 (6.3-23.4)	10.5 (3.1-17.2)
RF	6.25	22.0 (10.9-32.8)	18.2 (4.7-34.4)	11.4 (4.7-21.9)

Table 5.5: Mean classification accuracies and the range of performance among subjects on the vowel prediction task using the LOSO CV method.

		Classification accuracy (%) per feature set		
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	25	**53.62 (40.63-68.75)**	**45.10 (34.37-59.38)**	**46.80 (31.25-67.19)**
SVM	25	38.57 (25.0-51.56)	37.21 (28.13-48.44)	38.07 (21.88-50.0)
LR	25	37.14 (25.0-46.88)	35.87 (26.56-46.88)	38.14 (21.88-50.0)
KNN	25	46.16 (26.56-62.5)	42.68 (31.25-54.69)	41.83 (31.25-56.25)
DT	25	40.77 (31.25-62.5)	33.66 (18.75-45.31)	35.51 (23.44-54.69)
RF	25	45.17 (32.81-57.81)	37.93 (26.56-46.88)	39.49 (21.88-53.13)

Table 5.6: Mean classification accuracies and the range of performance among subjects on the level prediction task using the LOSO CV method.

		Classification accuracy (%) per feature set		
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	25	51.9 (32.8-78.1)	47.4 (29.7-78.1)	34.2 (23.4-50.0)
SVM	25	44.1 (29.7-60.9)	41.8 (28.1-54.7)	34.5 (23.4-46.9)
LR	25	42.0 (29.7-56.3)	41.6 (29.7-51.6)	34.3 (23.4-46.9)
KNN	25	**53.8 (35.9-75.0)**	**48.6 (32.8-71.9)**	**36.9 (28.1-43.8)**
DT	25	44.5 (29.7-59.4)	38.8 (26.6-50.0)	30.8 (23.4-40.6)
RF	25	51.2 (35.9-75.0)	43.9 (26.6-65.6)	32.1 (18.8-39.1)

Table 5.7: SVM classification accuracy on the vowel, level, and vowel-level prediction tasks, using the eFFR and sFFR time series data alone and in combination. Best performances are indicated in bold.

		Classification accuracy (%) per feature set		
Predictive task	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
Vowel & level	6.25	**50.9**	34.9	32.6
Vowel Only	25	71.3	43.9	**74.6**
Level Only	25	**61.6**	60.0	43.8

Table 5.8: SVM classification accuracy on the vowel, level, and vowel-level prediction tasks, using the eFFR and sFFR time series data alone and in combination with the addition of 'sex' as a feature. Best performances are indicated in bold.

		Classification accuracy (%) per feature set		
Predictive task	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
Vowel & level	6.25	**50.6**	34.7	32.2
Vowel Only	25	71.2	43.9	**74.7**
Level Only	25	**61.5**	56.4	43.8

addition of sex as a feature tends to result in slightly worse results.

The confusion matrix for the vowel-level class predictions on the eFFR and sFFR time series features is shown in Fig. 5.7. The majority of the predictions clearly lie on or very near to the diagonal, and 72.0% of the confusions occur within the correct vowel class and within ±10 dB of the actual level class.

The corresponding confusion matrix for the data trained and tested on the sFFR time series data is shown in Fig. 5.8. Although most of the heat is still concentrated about the diagonal, it is much more dispersed than in Fig. 5.7 above, and indicates that although the sFFR features result in good within-vowel-class predictive performance, they result in much poorer level-class accuracy.

In contrast to the above, the confusion matrix obtained by using the eFFR time series features alone (Fig. 5.9) reveals relatively weak vowel classification accuracy as many of the predictions lie distant from the main diagonal. However, the heat occurring along the main diagonal and roughly along the diagonals parallel to it reveal that the eFFR features seem to predict the level-class with relatively high accuracy.

The vowel-class confusion matrices obtained from testing and training the SVMs on the time series data is shown in Fig. 5.10. It is quite evident that the use of the eFFR features alone (right panel) results in many erroneous predictions, whereas using the sFFR features results in few. The use of both the eFFR and sFFR data (left panel) still results in strong predictive performance, however it is less accurate than the sFFR-only case, which achieved a classification accuracy of 74.7%.

Figure 5.11 shows the confusion matrices for the level-class predictions resulting from the training and testing of the SVMs on both the eFFR and sFFR time series data in combination (left), the sFFR data alone (middle), and the eFFR data alone (right). The sFFR features on their own clearly result in many level-class confusions, and erroneous predictions off by ±30 dB are much more common in comparison to the eFFR-only and eFFR and sFFR-combined results. For instance, the sFFR-only matrix reveals a ~12 times greater incidence of predicting 85 dBA when the actual class label is 55 dBA. Predicting a class label of 55 dBA when the actual label is 85 dBA occurs ~10 times more frequently

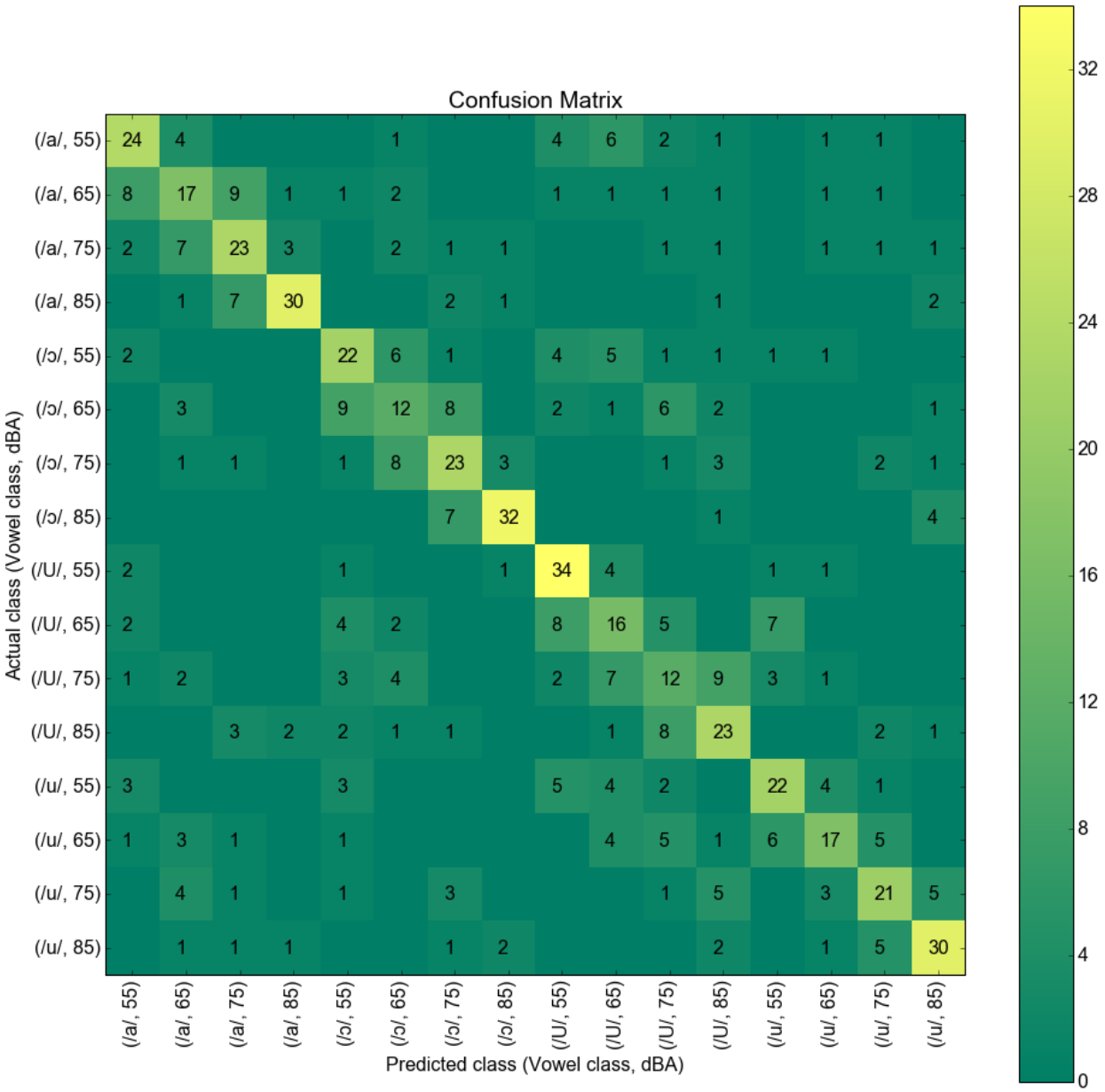

Figure 5.7: The confusion matrix corresponding to the SVMs performance on the vowel-level classification task using both eFFR & sFFR time series features, with the actual vowel and level class pairs indicated on the ordinate and the predicted class pairs on the abscissa.

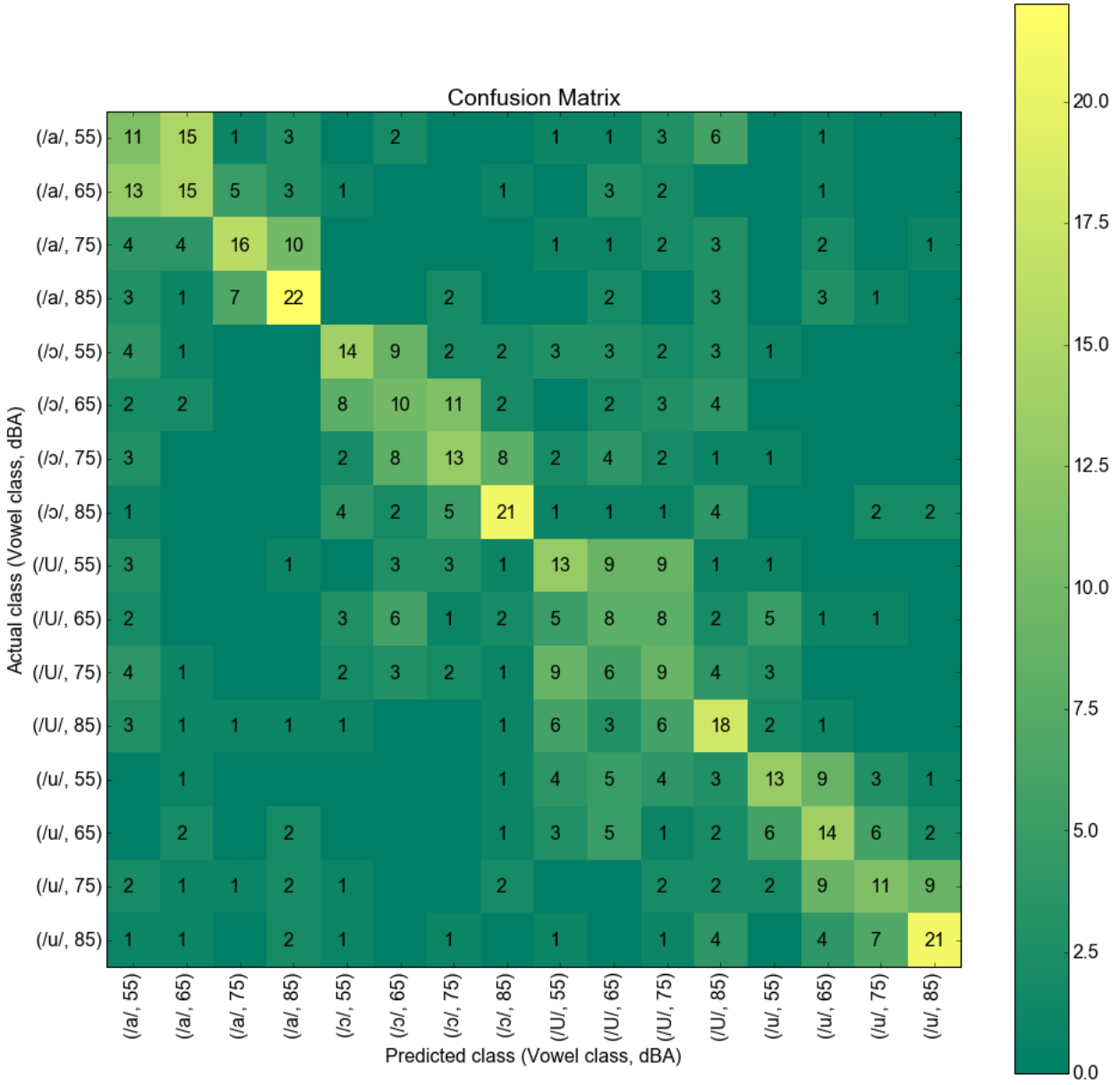

Figure 5.8: The confusion matrix corresponding to the SVM's performance on the vowel-level classification task using sFFR-only time series features.

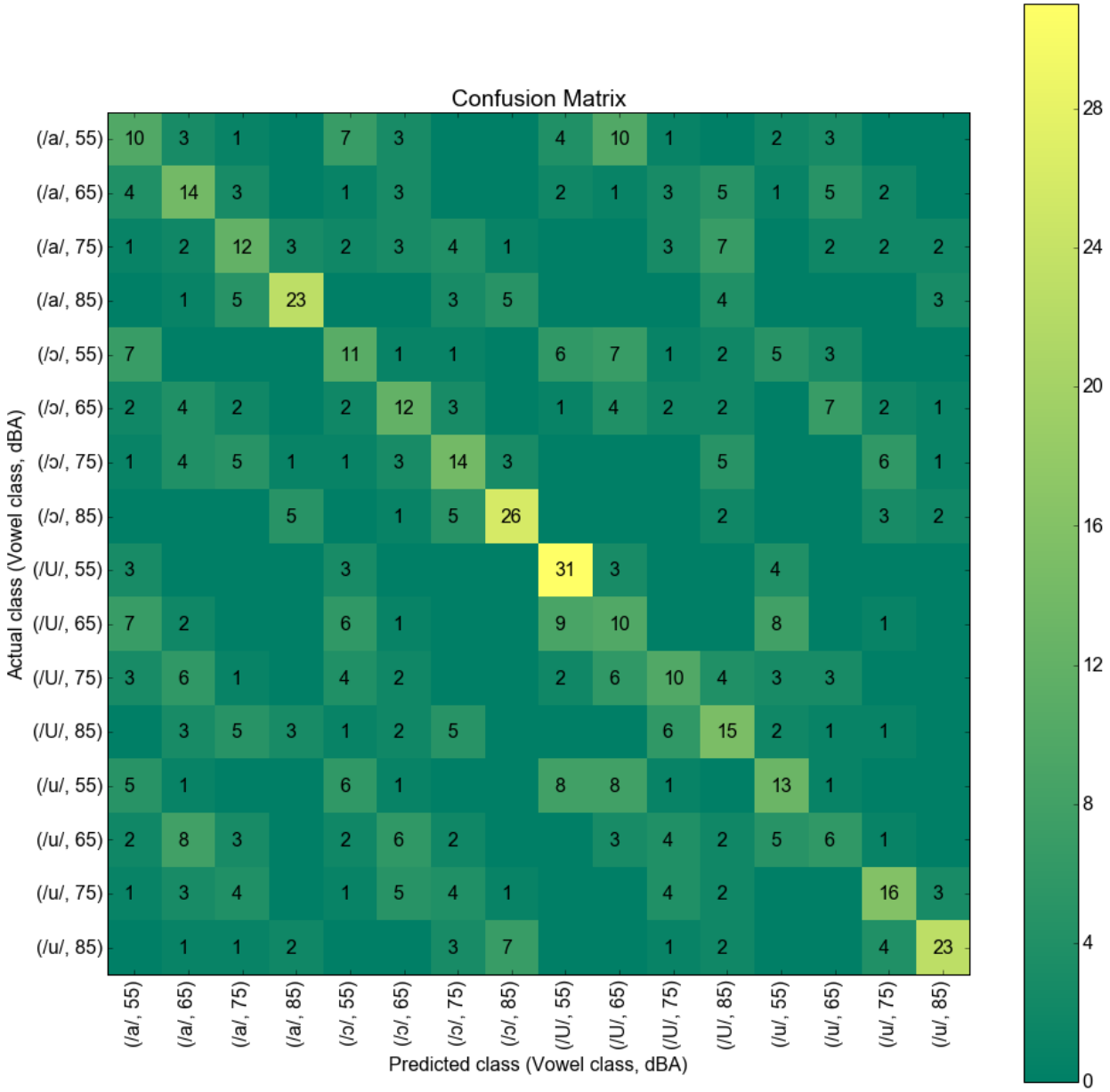

Figure 5.9: The confusion matrix corresponding to the SVMs performance on the vowel-level classification task using eFFR-only time series features.

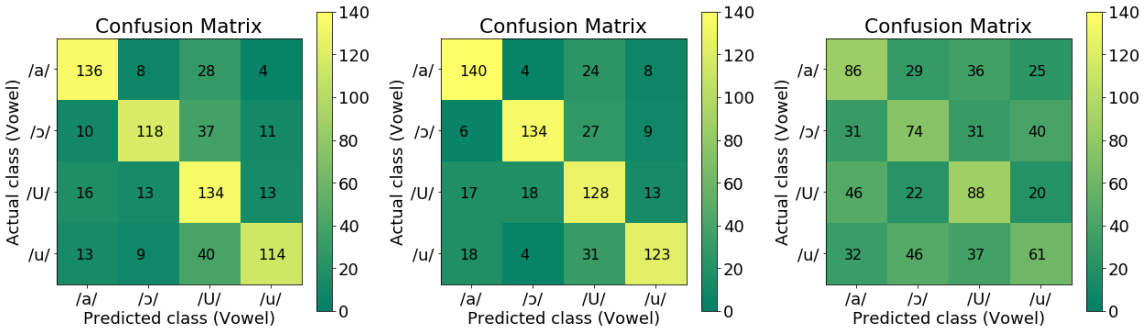

Figure 5.10: Confusion matrices for the SVMs performance on the vowel-class categorization task obtained by testing and training on the eFFR & sFFR features (left), the sFFR-only features (middle), and the eFFR-only features (right).

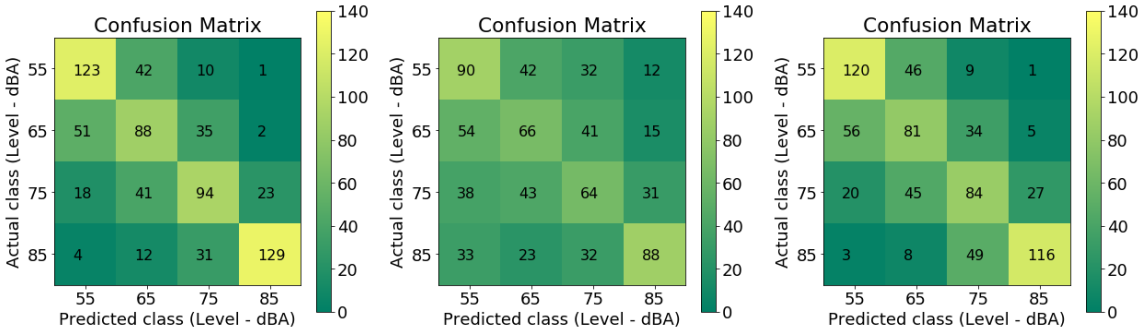

Figure 5.11: Confusion matrices for the SVMs performance on the level-class categorization task obtained using the eFFR & sFFR features (left), the sFFR-only features (middle), and the eFFR-only features (right).

than in the two competing cases as well. However, including the sFFR data with the eFFR does outperform the eFFR-only scenario, resulting in a classification accuracy of 61.5% and indicating that the sFFR waveforms do possess meaningful level-class related information beyond that contained in the eFFR. Here, 93.3% of level-class predictions are within ±10 dB of the actual level.

5.3.3 Signal similarity algorithms

The signal-similarity algorithms performed roughly the same, both in the intra-subject and between-subjects instances. The greatest classification accuracies were obtained using the intra-subject signal correlation algorithm, which attained 67.5% level-class accuracy (n=1) using the raw eFFR time series data and a vowel-class accuracy of 63.1% (n=1) using the filtered sFFR time series data. In general, vowel and level class prediction accuracies of ~55-65% were achieved throughout when using the eFFR time series data for level-class predictions and the sFFR time series data for vowel-class predictions. The highest accuracy scores attained for each algorithm in both within and between-subject instances are tabulated in Tables 5.9 and 5.10. Here, n is restricted to the range of 1-20, since no meaningful

Table 5.9: The best results obtained via the signal correlation algorithm – both intra-subjects and between-subjects – for each of the vowel, level, and vowel-level prediction tasks. The feature set used to obtain the results, along with the n-value corresponding to the best accuracy are indicated.

Signal correlation algorithm -best results across predictive tasks						
	Intra-subject			**Between-subjects**		
Label class	Vowel	Level	Vowel-Level	Vowel	Level	Vowel-Level
Accuracy (%)	63.1	67.5	38.1	60.8	55.8	26.8
Range (%)	18.8-90.6	46.9 - 90.6	12.5 - 68.8	43.75-78.1	31.25 - 78.1	6.25 - 59.4
Feature set	sFFR filtered	eFFR unfiltered	eFFR unfiltered	sFFR filtered	eFFR unfiltered	eFFR filtered
n-value	3	1	1	18	11	1

Table 5.10: The best results obtained via the signal subtraction algorithm – both intra-subjects and between-subjects – for each of the vowel, level, and vowel-level prediction tasks. The feature set used to obtain the results, along with the n-value corresponding to the best accuracy are indicated.

Signal subtraction algorithm - best results across predictive tasks						
	Intra-subject			**Between-subjects**		
Label class	Vowel	Level	Vowel-Level	Vowel	Level	Vowel-Level
Accuracy (%)	53.3	66.6	35.6	55.2	59.7	14.7
Range (%)	21.9-59.4	58.3-71.9	28.1-38.9	31.3-81.3	28.1-87.5	0-56.3
Feature set	sFFR filtered	eFFR unfiltered	eFFR unfiltered	sFFR filtered	eFFR unfiltered	eFFR filtered
n-value	1	1	1	3	3	2

performance increases were witnessed for larger n (see Fig. 5.12 for an example of the performance changes with increasing n-values). Plots showing the plateauing of performance and general decline for sufficiently large n are provided in Appendix H.

The greatest vowel-level classification results were obtained from the intra-subject predictions, which is not surprising given that the test sets are much smaller in size and that individual subjects have high test-retest correlations. Although not reported on here, tests using the amplitude spectra from both the eFFR and sFFR produced markedly lower best intra-subject results, typically in the range of 35-45% accuracy for both vowel and level class tasks despite their high test-retest correlations within subjects. The lack of phase-related cues may be a partial cause here, as the systematic latency changes observed in the eFFR waveforms are likely to positively influence the discriminative ability of both signal similarity measures.

As mentioned above, classification accuracy alone does not often convey the extent of a model's performance, and this can be seen in Figures 5.13 & 5.14, which both provide a clearer illustration of the level-class results obtained using the two signal-similarity algorithms. In addition to the obvious increase in accurate level-class predictions (i.e. '0 dB' differences), both plots show reveal that far fewer large errors occur among the predictions than would be expected due to chance, which results in more leptokurtic distributions (i.e. a sharpness in the peak of the distribution due to many scores being close to the mean).

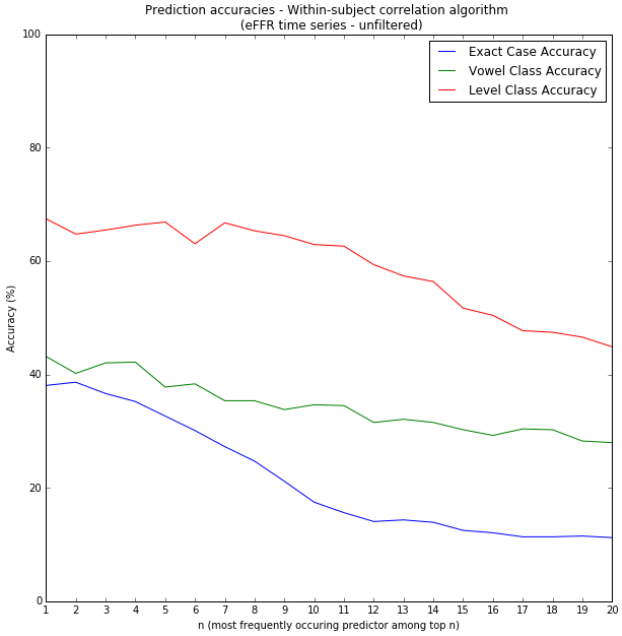

Figure 5.12: The predictive accuracy of the intra-subject signal correlation algorithm using the raw eFFR time series data is shown here as n increases from 1 to 20.

Generally, greater than 90% of the level predictions are within ±10 dB of the actual level.

5.4 Discussion

The classification results indicate that both vowel, level, and vowel-level class membership can be predicted with a good degree of accuracy using a variety of different classification methods and a variety of time and frequency domain features obtained from the speech-evoked FFR. Although a few studies have assessed vowel classification methods, to the best of our knowledge, these results represent the first attempt to classify both vowel and level. This is particularly important from an applied perspective, since the two most prominent issues with hearing aid performance have to do with lack of clarity or presence of distortion – which can result in speech misclassification – and loudness control, as the input level is often not subjectively experienced to be of appropriate or normal loudness once the signal has been amplified by the hearing aid (Anderson, 2017; Bidelman, 2017).

The discussion of the results is broken down into three parts, each one corresponding to the three different ML research sections presented above.

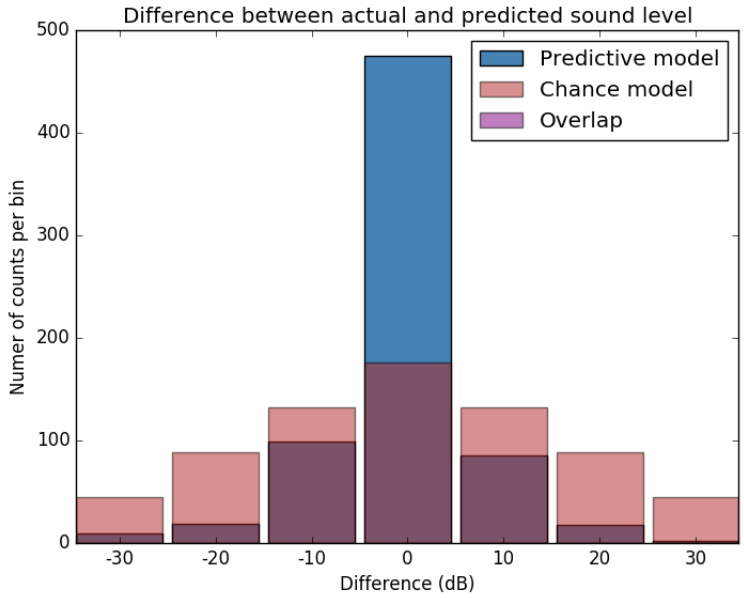

Figure 5.13: A bar plot illustrating the distribution of the differences between the actual and predicted sound level resulting from the best intra-subject results obtained (due to the signal correlation algorithm – see Table 5.9).

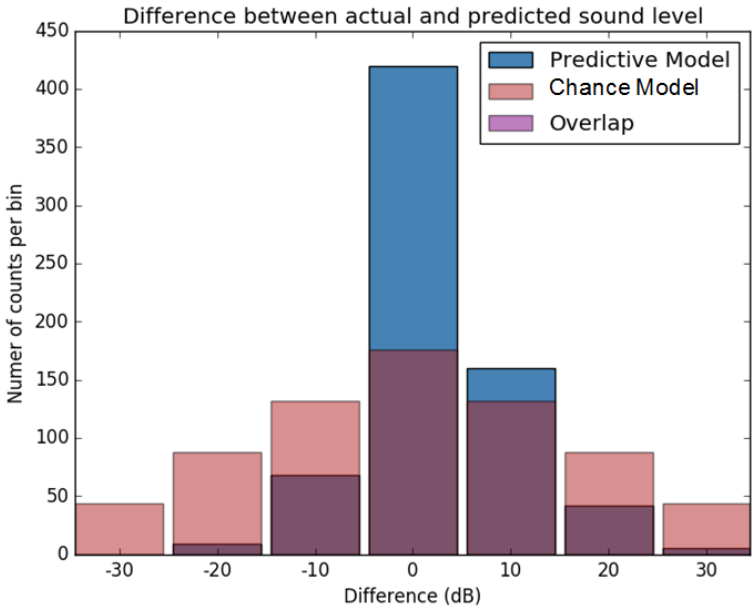

Figure 5.14: A bar plot illustrating the distribution of the differences between the actual and predicted sound level resulting from the best between-subject results obtained (due to the signal subtraction algorithm – see Table 5.10).

5.4.1 Harmonic amplitude spectra features

The amplitude spectra of F0 and its harmonics in both the eFFR and sFFR were obvious feature candidates, as Studies 1 & 2 (see Chapter 4) found statistically significant effects of level on a variety of subsets of this global feature set, and the vowel stimuli in Study 2 and the FFRs that they elicited differed in terms of their respective formants. This selection of features was further supported by the results of Sadeghian et al. (2011, 2012, 2015), as well as the work of Yi et al. (2017), Won et al. (2016) and Bidelman et al. (2013), as all these studies indicated that there was sufficient information contained in the amplitude spectra of the FFR for vowel classification purposes.

The classification of vowel-level class membership resulted in a maximum classification accuracy of 32.5%, obtained using an LDA classifier with both eFFR and sFFR harmonics as features. As is often the case, the reporting of classification accuracy as the sole metric of performance can appear to be quite underwhelming when there are a large number of classes. Although the LDA classifier performs 5.2 times better than chance, it is perhaps the confusion matrix (Fig. 5.2) that best represents the actual performance observed here, as a large proportion of the predictions identified the correct vowel class but missed on the level class, typically by ±10 dB. Whether refining the level steps in future studies would result in misses that are generally within the correct vowel class and nearer to the actual level target remains to be seen, but this amount of granularity alone might be sufficient for certain systems that seek to enhance clarity and minimize excessively unnatural loudness, for example.

Another clear trend in these results is that the vowel-class confusions that occur often correctly identify the level-class, which is particularly common for the 55 and 85 dBA cases (see Fig. 5.2). That both eFFR and sFFR features were required for optimal performance should not be overlooked, as the theory that a balance between both the envelope (ENV) and temporal fine structure (TFS) as represented by the eFFR and sFFR is crucial to normal speech comprehension and behavioural performance has been prominent in recent FFR literature (Ananthakrishnan et al., 2016; Anderson, 2017; Bidelman, 2016; Kraus & Anderson, 2012, 2014; Léger, Desloge, Braida, & Swaminathan, 2015). In the vowel-level categorization task, using the eFFR features alone tended to outperform the sole utilization of sFFR features in terms of classification accuracy regardless of the classifier employed (see Table 5.1 above). On the surface, this seems to indicate that envelope spectra cues provide better behavioural performance for both vowel and level classification than do temporal fine structure cues in clean speech (i.e. not noisy). Although this aligns with the fact that whispered speech can be well understood despite the TFS consisting entirely of noise

(Bidelman, 2016), the bulk of the results from this study reveal that the sFFR features yield better vowel identification than the eFFR features do (see Fig. 5.10, for example), which is the de facto metric addressed in such ENV/TFS studies. In general, there are substantial differences in ability between normal and impaired hearing populations with respect to the utilization of ENV and TFS cues (Léger, Reed, Desloge, Swaminathan, & Braida, 2015; Léger et al., 2015) and a great deal of further study is required here in order to shed light on their relative contributions in different populations and environments. Regardless, these results further indicate that the FFR might be useful in studying the effects of envelope and temporal fine structure cues in normal and impaired populations not only with respect to speech identification metrics, but also with respect to the dimensions of level and loudness.

The maximum vowel classification accuracy of 71.7% achieved on four vowels using an RF model on the amplitude spectra of the sFFR was less than the 83.3% accuracy obtained on five vowels by Sadeghian et al. (2011, 2012, 2015). However, there are several major differences between the two studies. Most notably, we achieved such accuracy across all four level conditions, whereas Sadeghian et al. (2011, 2012, 2015) recorded all data at a fixed 80.5 dB SPL. In this sense, the vowel classification performance achieved here is invariant to level, which is a known and important perceptual capability of the human auditory system (see Appendix A.4.5). Although not noted above, an SVM trained and tested on the eFFR time series data at the sole level of 85 dBA resulted in 80.8% accuracy (see Fig. H.1 in Appendix H), which is very close to the results of Sadeghian et al. (2011, 2012, 2015).

Another difference between the studies was that Sadeghian et al. (2011, 2012, 2015) used the first seven harmonics of F0 (i.e. F0-H7) as eFFR and sFFR steady-state features as well as four transient features to obtain their result, where here we have used sFFR harmonic features alone (F0-H13). Furthermore, their work captured six subaverages for each subject and each vowel, and utilized a LOO paradigm to train and test. With only 8 subjects (2F) in comparison to the 22 (11F) recorded here, and knowing the high within-session correlations that exist among normal-hearing subjects' FFRs, there is the possibility that model overfitting is biasing their results as there are 5 training points representing the exact test case in each trained model. However, that effects of level present in the data might produce smaller vowel-classification accuracies is not surprising, since nonlinear level-related growth patterns were clearly present amongst some of the harmonic features used as inputs here (see Section 4.3), and algorithms that seek to maximize the linear separability of feature space suffer whenever within-class variability is increased, all else being equal.

It is worth noting that since the LOSO training and test methodology employed here resulted in a noticeable decline in vowel classification performance, it is quite possible that the results achieved by the RF are themselves due to overfitting in some sense, as the 10-fold

CV models will tend to include a mix of all of the subjects and might have been learning decision trees that were overly fitted to specific responses. This indicates the need for further work, as there are likely to be several clusters within the normal-hearing population at large, evidenced by the fact that there are likely to be different listening modalities that influence the FFRs of normal-hearing subjects (Coffey, Colagrosso, et al., 2016), that small effects of sex have also been reported and that some people naturally possess absolute pitch. That clusters may be present is supported by the fact that KNN tends to perform well across both the 10-fold and LOSO CV paradigms, whereas DT and RF drop off in performance when the LOSO CV method is applied. So while LDA or KNN likely represent a more conservative evaluation of vowel categorization capabilities here when compared to the RF results, if a sufficiently large dataset was ever to be created, it seems likely that a KNN classifier or an ensemble model such as an RF would yield comparatively better results, or that some type of cluster analysis in the pre-processing pipeline and subsequent group/cluster membership prediction would enhance predictive performance related to vowel and level class.

The sFFR harmonic features appear to be more crucial for vowel classification in comparison to the eFFR features here, as the sFFR-only models outperformed the eFFR-only models in all instances, albeit only marginally in some cases. On the high performance end, however, the sFFR-only models far outperformed the eFFR models. But as previously mentioned, this might be biased by overfitting and so caution is warranted here.

Comparable level categorization performance was achieved among the LDA, KNN, and RF classifiers when using both the eFFR and sFFR harmonic features, with the LDA model yielding the maximum observed accuracy score of 52.6%. The eFFR-only results outperformed the sFFR-only results in all instances here, implying that the eFFR harmonics contain more level-bearing information than do the sFFR features. Although different classifiers tend to produce different types of level-class confusions in different proportions – a fact that should not be ignored when considering technological applications that seek to implement ML for level-related purposes – the 65 dBA case was disproportionately mistaken for the 55 dBA case throughout. The 65 and 75 dBA cases tended to be the most difficult to accurately predict as they were both frequently confused with levels within ± 10 dB of their actual level. This is in line with expectations produced by a simple visual inspection of the grand-average amplitude spectra and the corresponding harmonic growth plots across levels for all four vowels (see Section 4.3). Since nonlinear growth functions that tend to either saturate or decrease at high levels appear to be a consistent trend in the literature (see Section 2.3 and Appendix D) and in the results of Studies 1 & 2 above, it may be that the results of any level-related ML studies on FFRs may vary substantially depending on the sound levels used, and that certain classifiers might perform better if the FFRs are recorded

within certain sound level ranges (e.g. within the range where many features appear to have linear-like growth behaviour). Repeating this study using finer level steps (3-5 dB, perhaps) would shed light on this. Additionally, the use of more advanced ML techniques such as neural networks might enhance classification results across all sound level ranges owing to their ability to learn nonlinear functions (Witten et al., 2016).

That the results of the LOSO training and testing were very close to those obtained using 10-fold CV – aside from the RF vowel categorization result discussed above – suggests that the results may generalize fairly well across the population. However, there exists a wide range of performance across the subjects (see Tables 5.4 to 5.6). For example, the subject on the low-end only achieved a vowel-level classification accuracy of 14.1%, whereas the top performing subject achieved an accuracy of 50.0%. In the vowel categorization task, the subject classification accuracy ranged from 40.6% to 68.8%, and in the level task it ranged from 35.9% to 75.0%.

It may be that some subjects produce abnormal FFRs compared to the unimodal population means (i.e. they are outliers), regardless of the fact that they possess normal hearing abilities – or as previously mentioned – it may be that the subjects on the low end represent a cluster in the normal population that is underrepresented in our dataset, and so any model trained without some of their FFRs in the training set results in excessively low performance. One major caveat here is that no normalization of the harmonic amplitudes was performed, and that the large variation in FFR amplitudes may have resulted in poorer classification accuracies than might otherwise be seen had some type of normalization process been applied. Future work assessing the impacts of a variety of normalization techniques on this dataset and any future collections is recommended. As is almost always the case in ML, there are likely substantial gains to be had in optimizing the pre-processing and feature selection process here (e.g. using wavelets, a grid search to optimize filtering parameters, or a spectral technique such as the one used by Yi et al. (2017)). Again here, more advanced ML techniques that rely on clustering algorithms and neural networks might greatly enhance classification results, and warrant serious consideration for any future investigations along similar lines.

Finally, the addition of sex as a feature generally resulted in a slight decrease in performance across all classification tasks. However, since some small sex-level interactions were observed in Study 2, and since the literature supports the existence of such effects, the potential utility of sex as a feature in spectra-based classifiers should not be discounted in future studies.

5.4.2 SVMs on time series features

The results of using the eFFR and sFFR time series data to train and test SVM models may appear somewhat surprising at first glance given that it is often the case that using relatively large feature sets results in poor performance; however, by taking into account the very precise temporal encoding of the subcortical auditory system that is produced via phase-locking throughout the periphery and brainstem, and to the resulting, evident local minima and maxima that are especially visible in the eFFR grand-average waveforms (see Section 4.3), the results begin to make sense. Although the features of SVMs are not easily interpretable, the fact that effects of level were very consistently observed in Studies 1 and 2 with respect to eFFR major peak latencies (see Chapter 4), and that both the eFFR and sFFR waveforms were filtered and normalized here in ranges devised to increase the respective vowel class differences, it is not surprising that robust feature patterns capable of being exploited by the nonlinear Gaussian kernel emerged. It is worth noting that both steady-state and transient features may be selected via the SVM classifier, but there is no explicit selection of either performed on the user-end.

The vowel-level classification accuracy of 50.9% represents an 8.1-fold increase in performance over the chance accuracy of 6.25%. Interestingly, the eFFR and sFFR-only results were nearly identical here, but it bears keeping in mind that they are both pre-processed, and so any projection with respect to auditory system capabilities need be taken with caution. A vowel classification accuracy of 74.6% was achieved using the sFFR time series data alone, which far surpassed the 43.9% accuracy achieved using the eFFR-only data. A level classification accuracy of 61.6% was achieved using both the eFFR and sFFR data, although the eFFR-only model yielded 60.0% here, and so the gains resulting from the addition of the sFFR are marginal.

As a whole, these results again support the notion that eFFR features are relatively more important than sFFR features to level prediction tasks, and that the opposite is true of vowel prediction tasks. This is rendered particularly clear via the inspection of the confusion matrices presented in Figs. 5.10 and 5.11.

Although not reported here, exploratory SVMs trained and tested on all of the subaverages rather than on the combined subaverages from the test and retest conditions yielded very similar performance, which implies that applications that need to take recording time into account can likely get away with fewer sweeps (and hence shorter recording times) if the small sacrifices in accuracy are not an issue. Similar exploratory results indicated weaker performance on the unfiltered and non-normalized (i.e. raw) time series data, which is not surprising since the data are fairly noisy, especially in the lower frequency range, which can

mask some of the temporal features present.

Finally, the addition of sex as a feature once again had slightly negative impacts on classification performance, and does not appear to be a useful feature. However, since some small sex-level interactions and effects of sex were observed in Study 2, and since mean major peak eFFR latencies of females was observed to be less than those of males in all cases, it remains likely that sex would be a useful feature in time series based classifiers being run on much larger datasets that are more representative of the population.

5.4.3 Signal similarity algorithms

Both of the signal-similarity algorithms had comparable performance, ranging approximately from 58-68% for vowel and level classification accuracies, placing them roughly in the same range as the conventional ML algorithms discussed above. Unsurprisingly, restricting the dataset to intra-subject comparisons resulted in better performance than when the 'test' data from each subject was compared with the 'retest' data across all subjects. As shown in the results of Study 2 (see Section 4.3.2), most subjects had very high test-retest correlations among both the eFFR and sFFR amplitude spectra and the between their test-retest eFFR waveforms. The fact that the filtered FFR waveforms resulted in the highest vowel classification accuracy throughout, and that neither the eFFR or sFFR amplitude spectra performed as well as their temporal counterparts was unexpected, however.

Once again, the use of the eFFR resulted in the highest level classification accuracy, and the use of the sFFR time series data resulted in the highest vowel classification accuracy. Attempts to combine the test and retest subaverages as done with the time-series SVMs described above resulted in slightly weaker accuracies despite the boost in SNR. This is likely due to the fact that this averaging operation reduces the number of retest comparisons by a half, which is particularly costly when there are often a few highly similar waveforms in the retest set.

The bar plots showing the difference between the actual and predicted sound levels regardless of vowel class are of particular interest, not just in that they show a 2.4-to-2.7-fold increase in performance above chance accuracy, but because they reveal that ~92% of the predictions are within ±10 dB of the actual sound level, and because there is ~80% reduction in large misclassifications of ±20 dB. This could have implications for loudness control functionality in a BCI-HA implementation, as changes in the shape of the bar plot could indicate increased or decreased neural encoding of appropriate sound level, and perhaps loudness perception as a result.

Although these preliminary results are generally positive, the fact that all of the stimuli

in this study made use of a common F0 of 100 Hz should not be overlooked when considering the signal similarity results. Both of the algorithms described here would almost certainly suffer to predict either vowel or level class with any meaningful accuracy were the F0s of the stimuli to change across all 16 vowel-level classes. This does not limit their application, as it is of course possible to create a standardized stimulus set for clinical use or for hearing aid fitting for instance, and the application of other signal similarity algorithms such as dynamic time-warping (Huang & Jansen, 1985) might be able to account for such F0 changes. And if ever FFR recording become so ubiquitous as to enter into the 'big-data' arena, then such simple similarity algorithms might re-emerge as being useful without need of concerns for F0 frequency.

These results reinforce the notion that the FFR is a highly stable response that can be exploited as a result of this fact, and supports the idea that self-contained BCI technologies that do not require data from other subjects might be viable. Such voting style algorithms (e.g. n-largest or n-smallest) might be particularly well suited to applications that record individual responses, or that create averages over a small number of iterations, which is a highly desirable quality for applications where there are constraints of time, computational power, and data storage.

5.5 Conclusion

In summary, these machine-learning results provide a good indication of the type of performance that can be expected from features extracted from the FFRs of normal hearing adults in clean auditory environments. All mean classification accuracies exceed chance accuracy, and often to an extent that warrants the serious consideration of the development of prototypical systems and clinical applications designed to exploit this fact. However, substantial variation was observed depending on the classifier used, the classification task at hand, the cross-validation method employed, and the subject under consideration.

Future work

A variety of additional work is required in order to provide a more rigorous assessment of the performance limits of ML when applied to the FFR, and to better understand what is driving this performance. This includes:

1. The utilization of more complex stimuli and stimulus sets, including consonant-vowel pairs, vowels with normal pitch trajectories, speech sounds from both male and female adults and children, etc. in varying background conditions (e.g. white noise, multi-talker babble, etc.).

2. The recording and storing of all raw, unfiltered FFRs for offline filtering and sub-averaging. This would allow for the comparison of results between inputs with differing average SNRs, the utilization of phase-locking value as a feature, etc. In general, newer FFR recording systems store every trial by default for these same reasons.

3. The collecting of data from a larger number of subjects, and performing a more comprehensive battery of auditory and behavioural tests (e.g. SIN performance, collecting loudness ratings, assessing relevant speech-token confusions for comparison with FFRs, etc.), while recording more data for use as potential features (e.g. age, sex, race, musicianship, handedness, and so on.)

4. Exploring the use of more sophisticated ML techniques along the entire pipeline, including more advanced pre-processing, such as neural networks for adaptive noise cancellation (Gholami-Boroujeny, Fallatah, Heffernan, & Dajani, 2016), various normalization procedures, etc., feature selection and reduction methods (e.g. recursive feature elimination), and the use of more sophisticated classifiers, such as neural networks of hidden Markov models.

Chapter 6
On potential FFR-based BCI constructions for hearing aid optimization

6.1 Introduction

The research question addressed by this study was the following:

> *What are some potential avenues to pursue with respect to utilizing the FFR to inform the (automated) fitting of hearing aids? Which might be the best to pursue and why?*

The overview provided in Chapter 3 indicated that there is early but growing interest in such BCI-HA related constructions both by research and commercial institutions. Also provided were reasons as to why the speech-evoked FFR in particular might be a particularly strong candidate AEP – both relatively and absolutely (for exploration in this vein see Section 3.3). Although a system that makes use of all of the neuro-electrophysiological information available among the AEP responses would obviously be ideal – and such an integrated AEP-BCI system is likely to emerge at some point since it would enable to some degree an understanding of the signal preprocessing fidelity (i.e. the 'brainstem' responses) and the associated perceptual and behavioural outcomes (i.e. the cortical responses) (Bidelman, 2017) – this study limits its scope in order to highlight the potential utility of the FFR alone. It is our hope that one or more of the candidate constructions or some variation thereof might represent the best first step along the path towards the ultimate realization of integrated AEP-BCI system, and that the considerations provided here might help to guide or facilitate the initial designs produced towards such an end.

As previously indicated, the FFR is particularly ripe for exploration in this context at present. This is due to the prevailing view that perceptual and behavioural performance are generally reliant upon the availability of envelope (ENV) and temporal fine structure (TFS) cues and that these cues, whose presence or absence within the auditory system are observable via the FFR (predominantly via the eFFR and sFFR, respectively), are known to be represented and exploited to varying degrees in the sensorineural hearing loss population relative to the normal hearing (NH) population (Ananthakrishnan et al. (2016); Anderson (2017); Bidelman (2017)).

Our own results from Studies 1-2 (see Chapter 4) further support the notion that ENV and TFS cues known to be of importance for speaker identification and the conveyance of speech information are robustly encoded in the FFR within the conversational speech range in quiet, and provide additional and novel insights as to how the representations of both change with level in the time and frequency domains. Study 3 further indicated that these neurally generated representations can be exploited for automatic classification tasks with a degree of accuracy that is promising for BCI purposes. Importantly, these results indicated that in quiet conditions ENV cues present in the eFFR tend to be relatively more important for level classification purposes, whereas TFS cues present in the sFFR appear to be relatively more important for vowel identification purposes. However, both the ENV and TFS cues present in the FFR enabled vowel and level classification accuracies that consistently outperformed the chance model (often by a ~2-7 fold increase), which on the surface seems to add to the notion that both sets of cues contain sufficient information to be exploited for accurate perceptual processing in healthy auditory systems depending on their relative availability due to environmental conditions (e.g. white noise, reverb, multi-talker scenes, etc.) (Bidelman, 2017; Bidelman & Howell, 2016). Testing the correspondence between behavioural outcomes and their reliance on ENV and TFS cues in studies with carefully designed stimuli and the relevant ML classification accuracies on the FFRs obtained on these same stimuli would enable a better understanding of how well ML performance can be used as a proxy for behavioural performance here.

In general, the encoding of ENV and TFS cues in the auditory system as revealed via the FFR appears to be predictive of outcomes related to clarity as well as audibility and loudness (Anderson, 2017; Bidelman & Howell, 2016; Easwar et al., 2015b; Van Eeckhoutte et al., 2016), which are among the most problematic and basic performance issues faced by hearing aid users. Clearly ensuring audible detection of speech sounds or even discrimination between speech sounds is not sufficient from an audiological rehabilitation perspective, and the aim of any HA fitting should be to maximize comprehension and communication while maintaining comfortable loudness and good sound quality. Although these are obviously lofty goals, especially in light of the generally underwhelming performance of HAs at present (McCormack & Fortnum, 2013), converging evidence very clearly suggests that the FFR warrants exploration here in terms of its potential utility to aid in the fitting process. In particular, the FFR appears to be a good candidate measure for exploiting information related to processing in environments particularly difficult for HA wearers since it is a good predictor of SIN performance (Bidelman, 2017) and of speaker identification and streaming (Du, Kong, Wang, Wu, & Li, 2011). Taken as a whole, this is sufficiently promising to encourage what seems now to be the inevitable creation of at least some form of prototypical

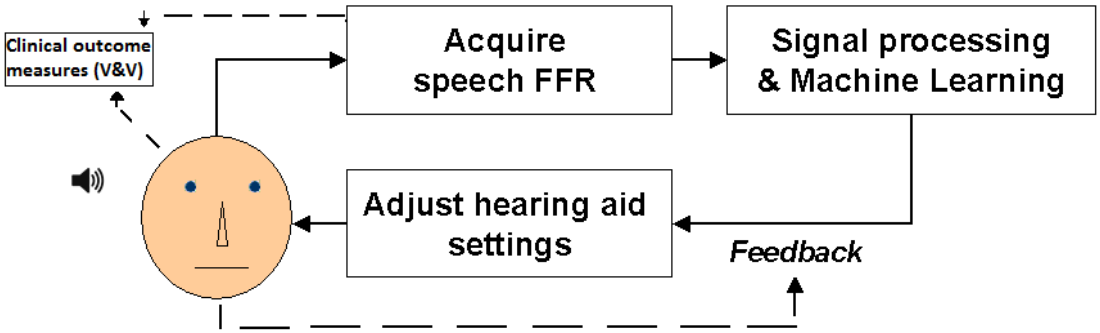

Figure 6.1: Schematic of a BCI-HA. Solid lines depict fundamental components; dashed lines represent potential human-based feedbacks (either clinician or user driven). User-based feedback could be realized via smartphone, for instance. Classical audiological outcome measures (i.e. verification and validation - or V&V) would allow the clinician to initialize the system appropriately, and could be used to interactively guide the system towards an optimal fitting.

BCI-HA technology in the near future.

This theoretical and exploratory study proceeds in three parts, beginning with a general description of the state and trajectory of the technological components in a BCI-HA system (Fig. 6.1) in order to establish technological feasibility, then proceeds with a discussion of the algorithmic approaches that are likely to be viable based on what is known at present, and terminates with a general discussion of the foreseeable stages of BCI-HA development and their associated hurdles. Although some work contemplating the general utility of the FFR as it relates to hearing aids has been done, no such study of the schematic and algorithmic space and the various related technological trends has been performed to date to the best of our knowledge.

6.2 BCI-HA components and trends

The analysis of the relevant technological trends can be subdivided into those pertaining to hearing aids and their components themselves, the inputs or stimuli, the BCI-related components, and a variety of feedbacks, as shown in Fig. 6.1. It should be pointed out that these abstracted components can be part of one integrated piece of technology (i.e. a fully contained self-fitting BCI-HA), or they can be distributed in a variety of ways. For instance, the sensors might be integrated into the HA itself with the signal processing and ML tasks being performed on a separate mobile phone or on an audiologist's laptop. Similarly, a separate FFR sensing headset might be developed that integrates the EEG electrodes with on-board signal processing and ML routines to be coupled with the HA at some prescribed timescale (e.g. daily, weekly, or monthly) depending on the use-case(s). Note that all of these approaches will require some amount of data logging, which is currently available, and allows for the collection and analysis of of all data related to user habits, device integrity, environmental conditions, and HA feature utilization (Camacho & Kuvadia, 2017). Finally,

although the intention here is to have the hearing aid settings changed automatically or to at the very least have a recommender system that analyses the logged data and provides clinicians / end-users with settings predicted to improve performance, it is not only possible but likely that there will be additional feedback provided by the end-users themselves (e.g. through a smartphone app that queries them regarding their performance ratings) and by clinicians as well as they perform validation and verification processes. Despite all of this, the signal acquisition and associated acoustic inputs, processing, ML and feedback pipeline will be taken to generally refer to the BCI component throughout, and the HA component can be taken to refer to systems lacking for any neurophysiological signal collection capabilities for sake of simplicity.

6.2.1 Hearing aids

Although there is an emerging trend towards self-fitting hearing-aids (see Section 3.1) and the component technology is rapidly evolving, Popelka and Moore (2016) point out that hearing aids will continue to be regulated, wearable medical devices and that they will also continue to be acoustic, battery operated, intended for continuous wear on a daily basis, and required to meet criteria related to comfort, cosmetics, battery life, and customizability with respect to frequency and level dependent amplification appropriate to the prescribed individual. They also indicate the need for reprogrammability, which is a requirement for BCI-HA systems. In fact, the facility of reprogramming via simple access to all relevant parameters through a well-designed application programming interface will be crucial, and could open the door to third party developers and/or the open-source community if this is deemed desirable from a business perspective.

The componential evolution mentioned above has quite fortunately kept pace with the ubiquitous trend towards miniaturization, and dramatic decreases in the size of microphones, receivers, processors, and batteries are being realized while performance capabilities are being maintained or improved upon (Dillon, 2012; Popelka & Moore, 2016). Although some current HA models make use of micro-electro-mechanical systems (Popelka & Moore, 2016), there is clear evidence of nano-electro-mechanical systems emerging in the near future (C. Wang & Madou, 2005), which will enable further design options (e.g. an array of directional microphones). Similar improvements in battery technology also continue to develop and are likely to improve dramatically in the coming years due to the reliance of an abundance of crucial technologies on their performance. Accompanying improvements in charging technology are also on the horizon, and rapid wireless charging is likely to have a great impact on hearing aid use and user satisfaction in the near future, as changing small batteries can be a particularly

difficult task for elderly persons who tend to possess diminished mobility and dexterity (Lu, Wang, Niyato, Kim, & Han, 2016). In order to address both the issue of battery drainage and the difficulty of frequent replacement, one novel approach is seeking to harvest energy from jaw movement (*Energy harvesting from jaw movement to power hearing aids*, 2014), and indicates the large amount of technological innovation occurring in this space.

Improvements in battery life and size are also likely to support improvements in processing power and in the digital signal processing needs that are becoming increasingly common in the latest generation of multi-channel hearing aids with a variety of settings for various environments and listening conditions (Popelka & Moore, 2016). Additionally, improvements in wireless technology has allowed for improvements in a variety of hearing aid functions (Mecklenburger & Groth, 2016), as they enable the utilization of external microphones and computational devices for additional digital signal processing functionalities, and allow for connectivity to cell phones, televisions, laptops, and more. Such seamless and ubiquitous connectivity is likely to result in additional gains, as big-data driven ML solutions will be able to exploit information such as geolocation data to train recommender systems for hearing aid settings, for instance (e.g. by realizing you are going to a busy restaurant or the opera, etc.) (Popelka & Moore, 2016).

Lastly, advances in 3D printing are likely to positively impact HAs as they will enable custom printing of components to ensure optimal molding and general fit. These advances will also enable preferential aesthetics, as visible components designed to match skin tone are easily envisioned, for instance. Since 3D printing of electronics at the nano-scale is very likely to emerge in the near-future (Flowers, Reyes, Ye, Kim, & Wiley, 2017; Shawrav et al., 2016), such advances are also likely to facilitate the creation and testing of truly personalized hearing aids that can be rapidly replaced or improved upon as new and improved schematics emerge (Edwards, 2007). Towards this end, Valentin, Viallet, and Voix (2017) have recently created a custom fitted in-ear wireless BCI. Generally, as the costs of these technologies come down, these advances might further result in improved access to appropriately advanced HA technology in the areas of the world that are presently most limited.

6.2.2 Signal acquisition considerations

Although the FFR can be accurately recorded using both MEG and EEG (Coffey, Herholz, et al., 2016), only EEG-based BCI-HAs are foreseeable at present due to the prohibitive size and cost of MEG systems. As is often the case, the performance of any BCI-HA will depend greatly on the quality of the data being captured and input into the pipeline that ultimately feeds back to alter the hearing aid settings. To this end, ensuring that adequate sensors are

available and capable of meeting the use requirements is critical.

At present, most FFRs are captured on a desktop or laptop computer via wired, wet electrodes common to the acquisition systems used in labs and clinical settings since they permit the attainment of sufficiently low electrical impedances required for detection. Outside of such settings however, wet electrodes are likely to be a grave cause for concern as they negatively impact the wearability and general ease-of-use of the technology. The recognition of this issue in the more general wearable sensor technology space has led to the development of wireless, dry, active (i.e. pre-amplified) EEG electrodes often made of novel materials (e.g. advanced polymers or metallic inks) with spatial configurations that enable efficient signal capture through hair (Brummer et al., 2013; Chen et al., 2014; P. Fiedler et al., 2016; Mathewson, Harrison, & Kizuk, 2017). Additionally, 3D printing of custom-fitted dry electrodes is now possible, which will further improve function, comfort, and aesthetics (Krachunov & Casson, 2016).

Of particular interest from a BCI-HA perspective is the recent emergence of both in-the-ear and behind-the-ear electrode systems with specific designs for compatibility with HAs and AEP capture (L. Fiedler et al., 2017; Kidmose, Looney, Jochumsen, & Mandic, 2013; Kidmose, Looney, & Mandic, 2012; Looney, Kidmose, & Mandic, 2014; Looney, Kidmose, Morrell, & Mandic, 2014; Strickland, 2016). This very clearly points towards the possible and likely integration of sensors with classical behind-the-ear and in-the-ear hearing aid designs, and that there is growing interest in potential BCI constructions that exploit AEPs (particularly focused on attentional correlates in these instances). Although such EEG systems are likely to meet the wearability requirements related to the discreetness and general aesthetic concerns of end-users and product designers, whether they are optimal for the types of BCI solutions envisioned here is another question entirely.

To date, most FFR montages have been simple single channel wired systems with a small set of reference and ground placements (Bidelman, 2015; Skoe & Kraus, 2010). However, recent advances in our understanding of the FFR generators (Coffey, Herholz, et al., 2016; Coffey et al., 2017; Tichko & Skoe, 2017) and a more general interest in multichannel montages has shown that electrode placement is likely to play an important role in the acquisition of FFRs (Bellier et al., 2015; Bidelman, 2015; Coffey, Colagrosso, et al., 2016), and different montages might be preferential depending on whether or not more peripheral or central pathologies are being treated.

One study by Bellier et al. (2015) is particularly germane to this issue in that it not only revealed that the SNRs of FFRs captured via a conventional 32-channel EEG system with a low sampling rate (5 kHz) more typical of clinical-grade systems were no different than those captured via a high-sampling rate system (13.3 kHz) more common of research-

grade systems. Additionally, their multichannel montage and natural 200 ms /ba/ stimulus allowed for the simultaneous recording of the FFR and CAEP and for the calculation of their respective topographies. The multichannel data further permitted the application of a denoising source separation algorithm, which significantly improve the SNR of the FFRs. Obviously a full-cap EEG system as they exist presently is not truly a wearable technology and their utilization is not likely to be as viable from a commercial perspective due to the difficulty of their use even among experienced researchers; however these results highlight the potential gains of a multi-channel system that captures activity from a broader covering of the scalp and that considerations in this dimension are certainly warranted. And of course, the larger multi-channel systems could still be useful in clinical settings where wearability (in the sense described here) is not of great concern.

Finally, although FFRs as captured in the laboratory often amount to very small files from a digital storage perspective, especially relative to the size of hard drives typical of desktop and laptop computers, it is easy to envision commercial BCI-HA systems that make use of storage contained with the HA itself, a small additional component or headset, or a smartphone or tablet. Since the latter are likely to have much more limited capacity in this dimension relative to laptops and desktop computers, considerations of storage options are obviously warranted. At present, there are a number of wireless smartphone coupling capabilities already being exploited by hearing aid companies and the most obvious solution for any sizeable data storage requirements is likely to be realized by relaying (pre-processed) information from the acquisition system to a smartphone, and then perhaps to the cloud if warranted and possible.

This type of BCI-HA system is likely to be of significant interest from a commercial perspective, since the availability of large amounts of high-quality data is often what limits the benefits of ML driven products. Enabling users to opt-in to anonymized data-sharing via the cloud, for instance, would allow both data scientists and research experts to mine the data and to refine their models for improved performance. For example: since significant inter-individual differences exist in the FFRs of normal hearing subjects and likely in the HI population too, clustering algorithms applied to a big data set tagged with metadata might enable custom algorithms to be developed for subjects who tend to have reduced F0 but large 2F0 (i.e. H2) representation in their eFFRs at higher levels but who perform as well behaviourally as subjects in their matched groups (e.g. age, sex, race, hearing loss profile, etc.) and who might otherwise be difficult to treat using F0-based algorithms devised from more commonly observed behaviour as a result. Such phone-to-cloud based storage functionality could further take advantage of more general health-tracking platforms such as Google Fit and Apple Health, which enable the tracking of health data from built-in

functions within the phones themselves (e.g. step counters and heart rate monitors), as well as third party applications and devices. Knock-on benefits such as the removal of pulse artefact and the investigation of a variety of comorbidities could possibly be realized as a result of such data integration as well.

Since signal processing and machine learning are both so ubiquitous in the modern world that it would require substantial efforts to go a day without being in some way influenced by their doings, it is no surprise that advances in these areas are being applied to HA technology wherever possible. Advances in signal processing related to noise reduction, beamforming, frequency shifting, automatic switching of listening mode based on predicted listening environment, etc. have all been realised to some degree in recent years (Dillon, 2012). Advances in the ML space have already led to a variety of attempts to create BCI-HAs referred to as 'cognitively controlled hearing aids' that exploit deep learning algorithms to isolate and amplify the stream being attended to by the HA wearer (Edwards, 2007; L. Fiedler et al., 2017; Popelka & Moore, 2016; D. Wang, 2017). Furthermore, the interaction between the two fields and its applicability to FFR-based BCI-HAs has already been demonstrated, as seen in work from our own lab (Gholami-Boroujeny et al., 2016) which showed that neural network-based adaptive noise cancellation can boost SNR and could therefore be used to reduce the overall recording time of FFRs when a specific SNR is desired. Additionally, studies highlighting the potential for automatic classification of FFRs have been performed (see Chapter 5.5 for a detailed discussion), and it has been shown that even single trial FFRs can be automatically classified with a decent degree of accuracy (Yi et al., 2017). Finally, work exploiting genetic algorithms has enabled the efficient searching of hearing aid parameter space in conjunction with user feedback in order to converge upon more optimal fittings than typically arrived at via conventional methods (Liang, Guo, Xi, Xie, & Zhao, 2017; Takagi & Ohsaki, 1999, 2007).

Looking forward, the same advances in speech recognition due to deep learning algorithms (Deng & Platt, 2014; D. Wang et al., 2016) will likely facilitate a variety of novel BCI-HA constructions, since smartphones and wearable microphones could perform a thorough scene analysis and accurate prediction of the speech content in each stream in near real-time, which would obviously be of great benefit for comparisons between auditory inputs and the neural markers of processing fidelity such as the FFR as well as perceptual classification (e.g. via the CAEP). The combination of all of these related advances will likely result in the development of expert systems that make use of cutting-edge signal processing and machine learning techniques not only for the automatic adjustment of gains and compression time constants, but for a variety of optimization targets along the fitting process from the initial assessment stage, to the selection of hearing aid(s), selection of various setting preferences

for music or telephony, verification and validation routines, and for real-time adjustments and automatic switching as well.

6.2.3 Feedback

A variety of feedbacks are possible in a BCI-HA system, the most basic of which refers to the feedback obtained from the online signal processing & ML routines, which will adjust hearing aid parameters in order to minimize some cost function (thereby maximizing certain targeted outcomes). Although a subset of these potential algorithms are discussed in detail below in Section 6.3, a simple example here could be an algorithm that adjusts the hearing aid gains and compression time constants in order to maximize the amplitude of F0 for a variety of vowel stimuli presented across a variety of sound levels. Note that although the extraction of F0 and its harmonics from responses to conversational speech would be ideal, early systems would presumably be built to exploit responses to several sweeps of pre-recorded vowels, either at device power up, or throughout the day, for example, in order to provide sufficient feedback for fitting purposes.

As mentioned above, it is also easy to envision systems that allow for the incorporation of user feedback in addition to the use of the FFR-based feedback, and such mixed systems might be able to learn individual preferences as they are reflected neurally over time (e.g. with respect to ENV and TFS cue representation). This is particularly important since an analysis of the preferential settings of users of self-fitting hearing aids in noisy rooms has shown great variability that is not well predicted by age, sex, or hearing loss (Nelson & VanTasell, 2016). As a result, it might be more desirable to first learn individual preferences using data logging (assuming their self-selected settings amount to verifiable performance gains along meaningful outcome measures) and to attempt to maintain FFRs similar in profile in similar environments rather than prescribing from population-based algorithmic approaches. Further evidence in favour of the self-directed feedback approach comes from the positive results in the interactive evolutionary computing domain, wherein genetic algorithms and user feedback about HA performance along a variety of dimensions have been used in order to evolve towards an optimal parameterization (Liang et al., 2017). Generally, incorporating user feedback is likely to expedite the time required to converge towards an optimal fitting, and to ensure that the user is satisfied with the performance of their device.

Feedback generated by clinicians is also foreseeable, especially in the earliest stage of BCI-HA development, as the earliest systems are likely to be administered by trained professionals in a clinical setting, and expert knowledge would allow for initial settings that are nearer to the user's optima than they would be if utilized out-of-the-box. Initial audiological

intervention would also allow for a host of benefits, including the setting of expectations, training on device and ear care, and in training the user to understand and manipulate the various settings of their HA. Further to this, FFRs could be routinely collected out of the clinic and monitored remotely (i.e. in a telehealth setting) by an audiologist for potential changes due to neuroplastic and/or behavioural adaptation, changes in hearing status, or potential issues with the hearing aid performance itself. As an example here, if F0 amplitude is observed to generally be decreasing over time for a variety of stereotypical stimuli and the user complains about decreased abilities to attend to a speaker in noisy situations, then the audiologist might target fittings (either manually or via algorithmic selection in the BCI) that seek to maximize F0. The assessment of relevant outcome measures could be similarly monitored for improvement, and recall to the clinic might be subject to automated thresholds on this F0 measure going forward.

Generally, feedback in a BCI-HA can be thought of as changes in HA parameterization based on at least one FFR-related outcome measure. Allowance for end-user and expert feedback from a clinician, although perhaps ultimately undesirable, will likely be necessary and of great benefit for many years to come, as a fully self-contained BCI-HA will surely require a great deal of technological maturation. This same sentiment has been paralleled in the self-fitting hearing aid space, which Keidser and Convery (2016) point out has yet to breach the threshold of any truly self-contained SFHAs.

6.2.4 Stimulus considerations

The stimuli used in BCI-HA systems should ultimately be reflective of the statistical properties of the end-users acoustic environment, but owing to present realities related to recording time, analysis techniques, and system prototyping requirements, initial constructions are likely to make use of simple yet relevant synthetic stimuli. Although a set of synthetic vowels such as those utilized in Studies 2 & 3 might provide for a good initial assessment of BCI-HA feasibility related to algorithms based on ENV and TFS cues known to be relevant to behavioural performance, they are not without their limitations. One obvious issue here is that the constant onset and offset of this type of stimulus presentation is not reflective of naturally flowing speech, and it will likely not engage the compression settings in a manner that reflects their performance in real-world scenarios.

In order to account for this issue, Easwar et al. (2015a, 2015b) and Easwar, Purcell, and Scollie (2012) have made use of a running synthetic speech stimulus with different amplitude modulations for different phonemes. Their results are very promising in that they clearly establish correlations between eFFR detection and amplitude and audibility, and

that the number and strength of the signals detected relate to sound level and behavioural performance abilities, and possibly recruitment and loudness as well. While the focus of their work was on statistical signal detection (effectively identical to ASSR methods) and enabled very short recording times, it is not generally known whether mere detection would be sufficiently robust for ML and related BCI-HA purposes as smaller components related to F0 harmonics and TFS might be necessary for certain measures. Stated differently, verifying for audibility alone is likely insufficient to ensure optimal HA performance (Anderson, 2017; Bidelman, 2017).

Another issue with speech is that its broadband nature precludes any clear determination of the carrier frequencies involved in the capture of envelope modulations. To this end, Aiken and Picton (2008) suggest that using a highpass filtered versions of the same stimulus with different cutoffs might be appropriate to gain a better sense of which components contribute to formant capture and which contribute to envelope encoding. This could be highly beneficial in that it would permit greater insight into whether or not gains in ENV cues are the result of decreases in TFS (or perhaps vice versa), which is aligned with the notion that "bigger (F0) is not better", as relative gains here might actually represent the swamping out of TFS as ENV cues are processed by fibers that would otherwise be representing formant-related harmonics (Kraus & Anderson, 2014).

Ultimately, we echo the view of Aiken and Picton (2008) that a sentence-lengthed stimulus might be most appropriate for auditory system processing as analysed via the FFR. The creation of such a stimulus with a large number and variety of phonemes that are manipulated to have various modulation frequencies (similar to the stimulus used by Easwar et al. (2015a, 2015b, 2012)) would allow for a broad assessment of the perceptual and behavioural capabilities the HA wearer along a variety of HA parameterizations, would permit comparison among a variety of populations and within a subject over time, and would be easily amenable to recordings and tests related to conditions involving noise and reverberation.

6.3 Proposed FFR-based algorithms for fitting optimization

6.3.1 Possible approaches

Each of the following potential algorithmic approaches is discussed with respect to both its potential utility and its possible limitations. The approaches as described represent a further elucidation of this space as proposed in our previous high-level overviews (Dajani et al., 2013; Heffernan, Giguère, & Dajani, 2017). They are considered here in greater depth as a result of insights gleaned from the results of Studies 1-3 and the most recent literature.

The approaches, discussed in detail below, are as follows:

1. The restoration of the FFR by comparison with norms from the normal hearing population.

2. Maximizing the correlation between the stimulus and the response (possible with both the eFFR and sFFR in both time and frequency domains – see Chapter 4 for more here).

3. Maximizing the correlation between a given response and another baseline response (e.g. between an FFR recorded in noise and one recorded in quiet for otherwise identical stimuli and level).

4. The restoration of normal balance between the ENV and TFS cues as indexed via the eFFR and sFFR.

5. Maximizing classification accuracy on some set of machine learning tasks with FFR-based features as inputs (e.g. maximizing the distance between different vowels in the FFR space).

Before discussing the approaches above, it is worth pointing out that there is likely some meaningful degree of overlap between them. For instance, it may very well be that the restoration of normal balance between ENV and TFS cues might amount to maximizing classification accuracy for vowel-level tasks, or that maximizing stimulus-response correlations might result in the restoration of normal FFRs to the extent that is possible.

The restoration of more normal FFR patterns

The restoration of normal synchronous activity in the auditory nerve and subsequent processing stations up to the cortex seems like an obvious goal, and investigations of how this might be achieved at the level of the auditory nerve in particular have already been performed (Sachs, Bruce, Miller, & Young, 2002). However, work by Giguère (1998) and Giguère and Smoorenburg (1999) has shown that no such return to normal is possible in a strict sense under amplification. So the question then becomes to what extent the FFR of HA users can be returned to normal. Here, the question of what constitutes normal becomes quite important, since results from Studies 1 & 2 (see Chapter 4) as well as from other recent work (Coffey, Colagrosso, et al., 2016; Won et al., 2016) have revealed substantial variation in the FFRs of normal hearing subjects. Whether additional factors are at play here, including untested issues related to hidden hearing loss for instance (Bharadwaj, Masud, Mehraei, Verhulst, & Shinn-Cunningham, 2015), remains to be studied in more detail.

Disregarding the need for further basic research in this domain in order to understand what a normal FFR truly looks like, a variety of different approaches can be taken in order to optimize HA parameterizations for more normal responses. For instance, a cost function related to the weighted correlations or rms-differences (similar to those used in Section 5.3.3) could be used to obtain an overall measure of similarity between the FFRs of the HA wearer and those of the normal reference population. By intelligently iterating through the parameter space via gradient descent or some similar technique, the settings that result in the minimum difference or maximum similarity could be arrived at. In this sense, the metric targets all of the response space, and not just one specific feature, such as F0 amplitude in eFFR frequency domain, say. Of course, some subset of features can also be selected for use in an otherwise identical algorithmic implementation.

Interestingly, a case study reported by Kraus and Anderson (2012) that studied two different settings on a single subject reported that the setting that maximized F0 was preferred (implied by their reference to it as the 'most favorable cABR (FFR)'). However, a closer inspection reveals that even though the favoured setting had substantially higher F0 representation, both settings yielded robust representations, while the preferred setting also clearly possessed greater amplitudes at the higher harmonics from H4-H6. Since a good deal of research supports the notion that these higher eFFR harmonics (i.e. ENV cues) are present in normal hearing subjects in quiet, but that they are significantly reduced or totally absent in difficult listening conditions such as in noisy or reverberant settings (Al Osman et al., 2018; Bidelman, 2017; Laroche et al., 2013; Prévost et al., 2013), it may very well be that the presence of these higher harmonic ENV cues typical of normal hearing subjects are what account for the HA users preferred setting in this instance.

Generally, while it may be the case that the optimal fitting results in the restoration of more normal FFRs, it is likely that this will depend on the extent and severity of the hearing loss, since some subjects with profound impairments might require the use of strategies not employed in the auditory systems of normal subjects (e.g. frequency shifting / remapping algorithms) to achieve the best outcomes for their particular circumstance. Furthermore, as will be discussed more thoroughly below with respect to algorithmic attempts to balance ENV and TFS representations, there are known differences in the way in which aging and impaired auditory systems rely upon the various cues present in a stimulus (Anderson, 2017; Bidelman, 2017; Bidelman & Howell, 2016) and it may very well be that emphasizing the TFS representations via the sFFR at the cost of the ENV representations via the eFFR could constitute a best-fitting in many cases. There may also be issues here pertaining to the length of time someone has possessed their impairment and whether or not they have been using amplification and for what duration, etc., as some subjects may have altered

their overall processing functions to a point beyond recovery via whatever plasticity remains. Therefore, while this approach might work well for subjects, such as those with a recent and mild impairment typical of age-related hearing loss and should certainly be considered as an algorithmic approach going forward, there is good reason to believe that it will not function as a catch-all in this space.

Maximizing the correlation between stimulus and response

The maximization of the stimulus-response correlation was first proposed by Kraus and Anderson (2012), and might be a simple yet effective approach towards utilizing the FFR in a BCI-HA. Although there is no explicit mention of which representation might best be used, it seems clear from the context that they are referring to the correlation between the stimulus and the eFFR. Their stated aim here is to maximize the representation of the temporal and spectral cues of the speech signal in the eFFR. While a good correlation here is likely a good indicator of such a neural representation of these cues, we have previously indicated that this might not be the best indicator of the quality of the internal representation since there are well-known transformations to the stimulus that occur all along the processing chain from the ear through to the cortex that result in a variety of nonlinearities (e.g. distortion products, stochastic resonance, suppression, etc.) that are consistent with normal auditory function.

Putting aside any consideration of such transformations, a greater elucidation of what is meant by stimulus-response correlation is required here in our view. Specifically, correlations between an appropriately lowpass filtered and resampled stimulus (in order to match the sampling rates of the FFR and the stimulus) are most appropriate. Furthermore, eFFR correlations should likely be performed with respect to the Hilbert transform of the stimulus (see Section 4.2), and the utility of correlations in both the time and frequency domains for both the eFFR and sFFR representations should be examined.

Maximize correlation between a response and another baseline response

As mentioned above, known transformations that occur in the auditory system impact the FFR and differentiate from being a simple, high-fidelity neural representation of the lowpass filtered stimulus. Given this, optimizing HA parameterizations by using a simple cost-function designed to assess response-response correlations might be preferred. If the baseline response here is chosen to be the response of the normal hearing population, then this case devolves to that analyzed in (1) above. However, there are some unique opportunities here that might make use of intra-subject responses or perhaps even the preferred responses of

an appropriately matched hearing loss population, say. For example, Anderson, Parbery-Clark, Yi, and Kraus (2011) found that older adults who performed better on a hearing-in-noise test had larger correlations between their FFRs in quiet and their FFRs obtained in six-talker babble noise (they also found that the higher performers had relatively larger representations of F0 than the poorer performers). It may therefore turn out that algorithms that seek to maximize the correlation within a subject between aided FFRs obtained in quiet when behavioural outcomes are favourable and aided FFRs obtained in a variety of difficult tasks might result in an optimal fitting. That such an algorithm might rely solely upon the individual being fitted would certainly simplify the process and considerations related to data sharing, model updating, etc.

Restoring the balance ENV and TFS cues

The notion of restoring and maintaining the balance between ENV and TFS cues via their representations in the eFFR and sFFR has gained prominence recently (Ananthakrishnan et al., 2016; Bidelman, 2016; Kraus & Anderson, 2012, 2014). Although the evidence is mixed, it generally appears that TFS cues are of greatest importance with respect to speech intelligibility in the HI population. Based on the results of Studies 1-3 and the fact that many studies on the effects of noise and reverberation show that the eFFR harmonics above F0 (H2-H6, say) are often entirely absent in these adverse listening conditions despite being present in quiet, we hypothesize that the restoration of these harmonics in quiet might allow for better clarity and more appropriate audibility. However, since ENV and TFS are differentially effected by changes in level (Ananthakrishnan et al., 2016) and adverse listening conditions (Anderson, 2017; Bidelman, 2017), with F1 representation being particularly vulnerable in HI populations (Anderson, 2017), we are in agreement with Kraus and Anderson (2014) in that increases in ENV cue representation should not come at the expense of TFS encoding. In fact, we further hypothesize that in the absence of being able to restore the ENV/TFS balance, parameterizations favouring F1 representation will likely be of greatest value. The reason for this is that the evidence suggests that TFS cues are most important for spatial hearing and selective attention in older listeners and for speech intelligibility in reverberant conditions (Anderson, 2017).

In general, the evidence from FFR studies of the HI population seems to indicate that ENV cues are good indicators of sensation level or audibility (Ananthakrishnan et al., 2016; Easwar et al., 2015a, 2015b) and that they are still required for speaker identification, object formation, and streaming, but that TFS cues require more robust encoding for appropriate speech intelligibility or clarity (Bidelman, 2017). Since F0 representations do not degrade nearly as much as F1 representations do in challenging environments (with F0 potentially

increasing at certain noise level - see Prévost et al. (2013)), algorithms focusing less on identifying who is speaking and more on what they are saying might be particularly advantageous. Anecdotally, the recent proof-of-concept study by Bellier et al. (2015) on the recording of FFRs via speech transmitted wirelessly to a hearing aid worn by normal subjects showed that both HA settings that were tested (identical aside from gains) dramatically impacted the amplitude of both the eFFR and sFFR representations of F0 while greatly diminishing the amplitudes of the F1 related harmonics. This is consistent with the notion that hearing aids have tended to mainly increase audibility but don't necessarily improve clarity (Anderson, 2017). That the real-ear output of the HA at different settings has such a large F0 component compared to the stimulus delivered via insert earphones and that this F0 component dominates the sFFR might hint at the weakness of typical built-in HA settings, especially with respect to clarity considerations.

Considerations with respect to the balance of ENV and TFS cues as encoded in the FFR certainly require further investigation; however, from an applied perspective, it seems that a variety of initial strategies could be easily tested here and that they might be fruitful by comparison to conventional algorithms.

Maximizing classification accuracy in ML tasks using FFR-derived features

The maximization of the classification of a selected machine learning classifier trained and tested on FFR-derived features has been conjectured from the outset of this research endeavour. This approach has appeal since it will by some measure require maximizing the separation between classes in the neurally derived feature space to the extent possible given the range of variability within a subject. The specific task is of importance here, since the results of Study 3 indicated that there appear to be differential effects of eFFR and sFFR features on classification performance with respect to level, vowel, or vowel-level tasks. Furthermore, these same results indicate that the best choice of ML algorithm might change depending on whether or not some function of level or loudness is being optimized for, or whether speech intelligibility or clarity is the target, or perhaps some balance between the two.

Constraints would be of particular importance in such an algorithm, since optimization might otherwise result in maximally separable FFRs of minimal behavioural relevance to the end-user if all of the stimuli under consideration are dramatically shifted in frequency, say. This is quite simple in practice, of course, but warrants mention since the general optimization of any ML task might bare little correspondence with behavioural gains. Another caveat here is that although ML results trained and tested on FFRs obtained in quiet from normal hearing subjects have been promising, it is not known whether or not similar classification

accuracies will be obtainable in HI populations, especially in noisy or reverberant conditions.

Again, some basic research to establish whether such an approach is valid is necessary. On its surface, however, this approach might be particularly valuable for subjects with severe impairments, as it may result in optimal speech discrimination between phonetic classes assuming an appropriately selected stimulus set is used, and frequency-shifting HA technology has been available for well over a decade (Alnahwi, AlQudehy, et al., 2015). Indeed, a similar approach is often taken with cochlear implants, wherein neural firing is altered in order to separate out speech classes, rather than to attempt to restore normal firing patterns (Choi & Lee, 2012).

6.3.2 Discussion

Typical of any emerging technology, the path from a bench-based prototype to a fully integrated, self-contained BCI-HA is likely to be a long and winding road, and will surely involve several feedback cycles between the basic and applied research domains before any viable commercial product emerges to test the market response. What is certain is that the technological landscape is rapidly changing, and the hearing aid design space is likely to be impacted by changes not just with respect to hardware, but with respect to advances in machine learning and their applications. Advanced algorithms for better outcomes measures related to a fitting alone are not sufficient, however, and it is likely that solutions that make compromises between algorithmic impact and efficiency of implementation, ease-of-use (particularly with considerations for elderly populations), and aesthetics will be required before any uptake beyond the early-adopter phase is realized.

Any successful designs in this space will likely be beneficial to both end-users and clinicians, since they might allow for more oversight of behavioural outcome measures via remote monitoring with fewer time costs for both end-users and clinicians alike. They might also permit for the identification of novel solutions in difficult-to-treat populations such as infants and persons with cognitive impairments. Finally, effective integration of an FFR-based BCI into a self-fitting hearing aid might permit for even more effective audiological care for the large proportion of the HI population that does not have or cannot afford access to professional audiological services.

Although there is no way of knowing which type(s) of FFR-based algorithm(s) might succeed or fail absent further investigation, it is likely that the technological evolutionary landscape here will ultimately result in a hybrid style system that makes use of stimulus-response and response-response correlations, tradeoffs between ENV and TFS cues, and ML task classification accuracies for increasingly better fittings in time. As always, caution is

required as such novel approaches are trialed and it is important to emphasize that some degree of invariance is required for proper statistical learning of auditory objects and streams, and that constantly changing the neural responses to sounds for short-term performance gains might have costly implications in the longer term.

6.4 Conclusion

There now appears to be sufficient convergence and overlap between the domains of auditory neuroscience, audiology, hearing aid technology, machine learning, and brain-computer interfacing to facilitate the construction of some prototypical systems that seek to exploit the information contained with the speech-evoked frequency following response. Which type(s) of system(s) ultimately emerge and whether or not there is sufficient commercial appeal remains to be seen, but we believe that the algorithmic solutions presented here provide a solid basis for near-term exploration of BCI-HA feasibility. With the application of sensible outcome measures and precautions, BCI-HAs might facilitate a wider adoption of and greater satisfaction with hearing aids by the people who need them most.

Chapter 7
Conclusion

7.1 Major Findings

The results of the four studies conducted to provide a better elucidation into the effects of level on the representation of vowels via the frequency following response, the accuracy with which these responses might be automatically categorized, and the potential utility of the insights gleaned in the process as they pertain to the construction of brain-computer interfaces for hearing aid fitting indicate that the FFR provides a sufficiently robust indexing of the encoding of relevant speech information to warrant additional pure and applied research in all of the aforementioned domains. The most significant original findings produced herein are summarized below.

The first four findings relate to research questions 1 and 2 (see Section 1.3), which asked how the representation of long and short vowels changed with level in the FFR of normal hearing adults, respectively. Question 2 further asked whether these representations differ between sexes, and if they are stable across test and retest conditions. Findings 5 and 6 address research question 3, which asked whether or not machine learning methods can discriminate between FFR responses to different vowels and/or levels with good accuracy. Finally, finding 7 addresses research question 4, which asked what potential avenues might be pursued with respect to utilizing the FFR to inform the automated fitting of hearing aids, and which might be the best to pursue and why.

1. The FFR to vowel stimuli presented at levels typical of the speech range provide robust and differentiable representations of both envelope and temporal fine structure cues present in the stimuli via the eFFR and sFFR in both the time and frequency domains. Previously published speech-evoked FFR studies of level effects, of which there are very few, have tended to focus exclusively on one representational domain (e.g. the amplitude spectra of the sFFR (Krishnan, 1999, 2002) or the analysis of eFFR waveforms in the time domain (Akhoun et al., 2008)).

2. The growth of eFFR representations of F0 is generally not monotonic-increasing with level in normal hearing subjects in quiet. This has never been previously reported to our knowledge, and has significant implications for studies that draw inferences

or create comparisons from results obtained at one fixed level. Particularly, since a great deal of the FFR literature uses stimuli presented near to 80 dB SPL, it may be that many reported findings do not hold for either different levels or different stimuli, and that further investigation and re-interpretation of a variety of results that use F0 amplitude as a predictor is necessary.

3. The growth of the eFFR harmonics (H2-H6) were the most consistent statistical indicators of level-related effects. The growth of the sFFR harmonics related to the first and second formants (F1 and F2) were also consistent indicators of level effects.

4. Although small effects of sex are apparent in the amplitude spectra and major waveform peaks of speech-evoked FFRs, very few sex-level interactions appear to exist. Consistent with previous literature on ABRs and FFRs, females tend to have larger amplitude spectra and shorter latencies than males within a given vowel-level condition. Due to the small effect however, no gains were made by the addition of sex as a feature with respect to classification accuracy in any of the machine learning tasks.

5. A variety of common classification algorithms are capable of achieving meaningful classification accuracies for both vowel, level, and vowel-level classification tasks trained and tested on features extracted from the FFRs of normal hearing subjects in quiet. Although some previously published work has yielded positive results for vowels presented at a single fixed level, no previous work has sought to classify FFRs based on effects of level to our knowledge. The ability to predict both vowel and level classes has important, positive implications for BCI based approaches to hearing aids that make use of automatic classification of FFRs since controlling for audibility and loudness in addition to clarity are all important considerations in an appropriate fitting.

6. On balance, the eFFR appears to provide more robust information related to presentation level classification accuracy than the sFFR. The sFFR appears to provide more robust information than the eFFR for the accurate classification of vowels.

7. Converging evidence based on the recent literature and the novel results above suggest that a variety of algorithmic approaches to FFR-based BCI-HA fitting are worthy of exploration. Since aided FFR recording has been independently validated by several laboratories using a variety of stimuli, and since a number of signal processing and machine learning techniques have been shown to be generally valid for use in this context, there are no known major obstacles to the testing of these approaches at present. The approaches discussed in Chapter 6 are good candidates for future exploration in this regard.

7.2 Limitations

There are several limitations to the work presented here. One such limitation relates to the fact that only synthetic speech stimuli were utilized, and they all possessed an identical fundamental frequency at 100 Hz, which is typical of an adult male speaker. Ideally, responses to natural stimuli across the range of frequencies typical of human speakers would be captured and analyzed in order to gain a more general understanding of the effects of level on the FFR and of the ability to use machine learning algorithms to accurately categorize said responses.

The sound delivery system utilized throughout this work is a limitation in the respect that the shielded insert earphones used cannot be viewed as good proxies to hearing aids, and so the FFR-related effects observed here may not be representative of those observed via aided delivery. Future studies related to BCI-HAs will need to replace these inserts with HAs or an in-ear hearing device of some sort in order to obtain a more realistic assessment of FFRs to amplified stimuli. This will likely require care when dealing with artifacts, processing latencies, etc.

Another limitation is that the recording device did not permit the capture of FFRs to individual sweeps, and only allowed for the collection of averaged responses. The recording of individual responses would allow for a better preliminary analysis of the types of BCIs that might ultimately be constructed. For instance, the ability to accurately classify a stimulus and its level using individual sweeps might suggest that real-time adjustments based on the FFR are possible, whereas requiring hundreds or thousands of sweeps before a useful signal can be exploited for fitting purposes would likely require scheduled fitting sessions. The best way to assess this is to record all of the individual responses, and to compose averaged responses as required, which was not possible here.

The FFRs upper limit of observable phase-locking is also a limitation, in that it hinders the ability to assess the effects of level across the vowel space more generally, since many vowels have second (and higher) formants that are not represented via the FFR.

Finally, these results only pertain to normal hearing, healthy young adults, and they may not generalize well to other populations. Moreover, we did not control for a variety of physical or sociocultural factors, including musicality, handedness, ethnicity, or bilingualism / multilingualism. Since these studies were undertaken with the broader goal of informing the design of therapeutic devices for the hearing impaired population, it would be ideal to have results from well-matched normal and impaired populations for comparative analyses. This would also permit a more direct assessment of the viability of various BCI-HA strategies.

7.3 Future Work

Based on the results presented here, we propose the following future work:

1. Repeating the FFR studies with a full-EEG montage and a system capable of recording single trials.

2. Recording responses from ~60-95 dBA in finer level steps (3 dB steps, say) in order to gain a better sense of the level-related growth functions of the eFFR and sFFR in the time and frequency domains.

3. Utilizing more sophisticated and comprehensive testing of auditory abilities, such as tests of musicality, listening mode, hidden-hearing loss, speech-in-noise testing, etc.

4. Acquiring loudness assessments from all subjects for all of the stimulus-level conditions in order to allow for level-loudness analysis.

5. Performing a similar study on hearing impaired populations in both aided and non-aided conditions, and including assessments of sensation level and loudness. Extensions related to the effects of reverberation should also be considered here.

6. Consider including assessments of context and transients using consonant-vowel stimuli (e.g. /da/, /ga/, /ba/, /du/, /gu/, /bu/, etc.). The representation of natural vowels with varying pitch and pitch tracks typical of males, females, and children would also be of interest here.

7. The utilization of more complex stimuli and stimulus sets, including consonant-vowel pairs, vowels with normal pitch trajectories, speech sounds from both male and female adults and children, etc. in varying background conditions (e.g. white noise, multi-talker babble, etc.).

8. The recording and storing of all raw, unfiltered FFRs for offline filtering and sub-averaging. This would allow for the comparison of results between inputs with differing average SNRs, the utilization of phase-locking value as a feature, etc. In general, newer FFR recording systems store every trial by default for these same reasons.

9. The collecting of data from a larger number of subjects, and performing a more comprehensive battery of auditory and behavioural tests (e.g. SIN performance, collecting loudness ratings, assessing relevant speech-token confusions for comparison with FFRs, etc.), while recording more data for use as potential features (e.g. age, sex, race, musicianship, handedness, and so on).

10. Exploring the use of more sophisticated ML techniques along the entire pipeline, including more advanced pre-processing, such as neural networks for adaptive noise cancellation (Gholami-Boroujeny et al., 2016), various normalization procedures, etc., feature selection and reduction methods (e.g. recursive feature elimination), and the use of more sophisticated classifiers, such as neural networks of hidden Markov models.

11. Evaluating all of the BCI-HA strategies presented in Chapter 6 in order to determine their viability.

7.4 Closing Words

It is our hope that this work contributes to the general body of knowledge in the fields of auditory neuroscience, audiology, and brain-computer interfacing for hearing aid applications, and that it helps guide further research and development in these domains.

Appendix A
An overview of the auditory system

The schematic shown in Fig. A.1 depicts a systems-level overview of the flow of an acoustic signal from its source in the external world all the way to the consciously experienced phenomenal state that is consequently induced within the subject (indeed, this schematic is illustrative of the path taken by all phenomena that excite one or more human sensory modalities). A brief glimpse at the schematic reveals that many complex structures and functions are involved in the processing path between the external event and its internalized representation by the subject, thus rightly imparting the notion that a great deal must be known about each and every of the (block-indicated) complementary aspects of auditory processing in order to fully characterize the system given a type or a class of input. Of course, there exist many gaps in our current knowledge of the auditory system's processing of speech in general, and this certainly holds true as concerns the encoding of speech sound level, which largely informs the subjective experience of loudness, an understanding of which is of crucial importance in the management of all hearing assistive technologies.

A.1 Life outside – the external world: auditory objects, events, and the basics of the acoustic signal

The external world referenced in the above schematic, which is more commonly referred to as the auditory environment or simply as the environment, typically consists of a vast array of auditory objects and events that reach our ears via the propagation of acoustic waves through a medium (usually air). Due to the massive number of combinations and permutations of sound waves that we as humans encounter and process, and due to the fact that these sound waves very often result in significant modifications to our behaviour as a result of this processing, it makes sense that our auditory systems must exploit the quasi-stationary assemblages of auditory qualities that we are able to perceive (and whose physical correlates are present in the waveforms themselves) in order to intelligently categorize sounds with respect to their sources and their meanings. The main qualities of pitch, timbre, location, loudness and duration are depicted in Fig. A.2 along with an understanding of their relation to both the acoustic waveform and the basic auditory processing undertaken in order to make sense of the auditory objects and events that transpire in our daily lives.

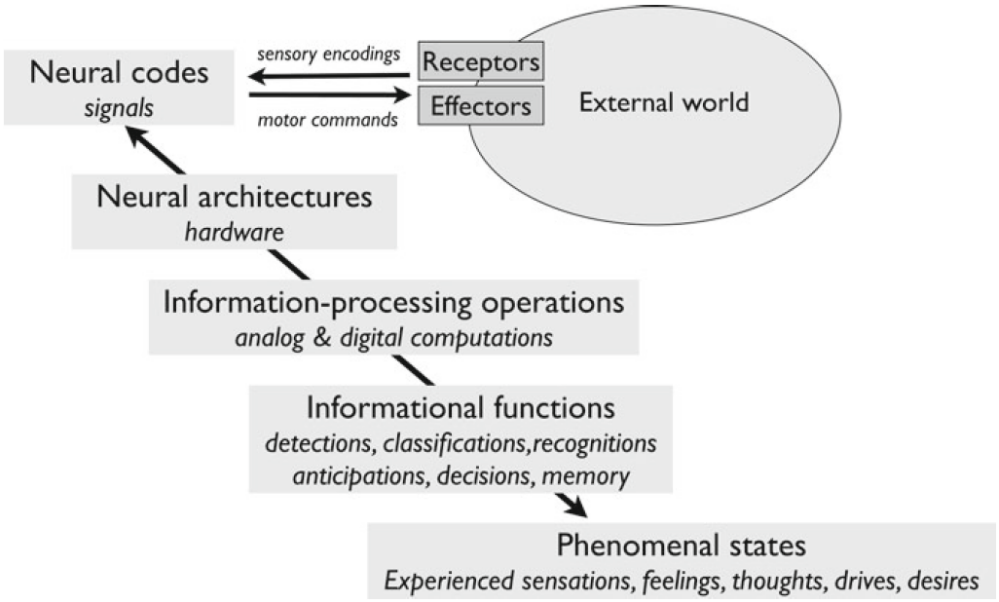

Figure A.1: A systems-level overview of the auditory system. An in-depth understanding of each subcomponent from the acoustic signal and its environment to the phenomenal states experienced by a subject is required in order to attain the goal of a full theory of audition. Reprinted with permission from "Toward a Theory of Information Processing in Auditory Cortex", by P. Cariani and C. Micheyl. In The Human Auditory Cortex (p. 352), by D. Poeppel et al. (Eds.), 2012, New York, NY: Springer New York. Copyright 2012 by Springer Science+Business Media LLC.

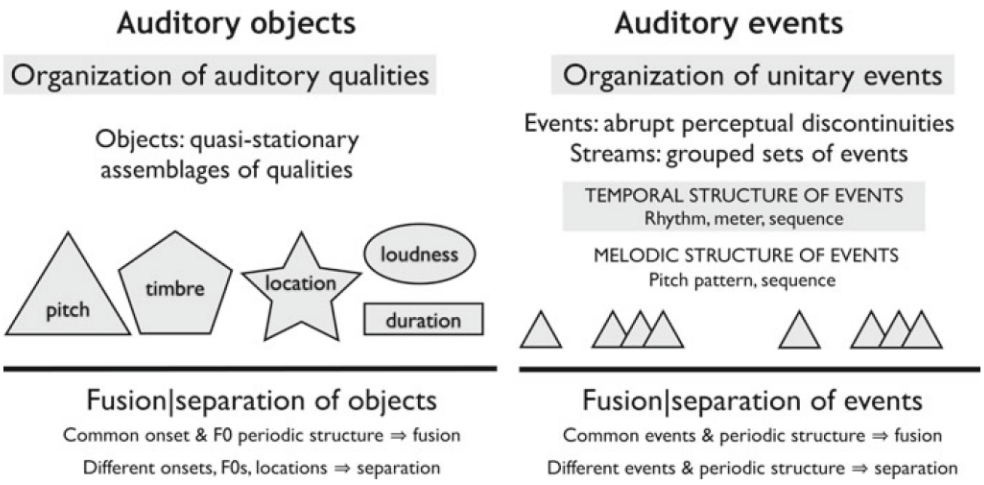

Figure A.2: Auditory objects and event. LEFT: The auditory qualities and factors that govern the fusion or separation of objects. RIGHT: The creation of auditory events and streams and the factors that govern their fusion and separation. Reprinted with permission from "Toward a Theory of Information Processing in Auditory Cortex", by P. Cariani and C. Micheyl. In The Human Auditory Cortex (p. 355), by D. Poeppel et al. (Eds.), 2012, New York, NY: Springer New York. Copyright 2012 by Springer Science+Business Media LLC.

Given the great complexity of the typical auditory environment, much of the research in audiology and auditory neuroscience has taken to characterizing the auditory system's response to what are by all intents very simple acoustic stimuli. Indeed, the most typical audiometric testing involves subjecting a listener to pure tone stimuli in the typical range of speech-relevant frequencies at various levels in order to determine the listener's threshold of hearing at those specific frequencies. These very same pure tones, as well as clicks and chirps, are frequently used in the characterization of various neuronal fibres and populations in auditory neuroscience experiments, and are also commonly employed stimuli in auditory evoked potential (AEP)related studies – including ABR and FFR studies.

Although pure (sinusoidal) tones do not occur in the natural world absent of manmade technology, they are nonetheless quite instructive models for developing an understanding of the acoustic signals that do comprise our environment. Typical pedagogical practices make use of this model to first introduce the notion of amplitude in order to lead into the concepts of sound pressure, intensity, and loudness as they apply to more complex stimuli. I will define these latter concepts and their respective levels for the sake of clarity, as they are often used interchangeably and often erroneously as a result (Yost, 2013, see) for more details relating to these concepts).

*A **Decibel (dB)** is a logarithmic unit used to express the ratio of two values of a physical quantity. Most frequently, "one of these values is taken to be a standard reference value, in which case the decibel is used to express the level of the other value relative to this reference".*

***Sound pressure** is the local pressure deviation from the equilibrium atmospheric pressure that is caused by a sound wave. The SI unit of sound pressure is the pascal (Pa).*

***Sound intensity** is a measure of sound power (i.e. the rate at which work is done) per unit area in situations involving sound. In most contexts, the measure is assumed to be taken in the air at the listener's location. The SI unit for sound intensity is the watt per square meter (W/m^2).*

***Sound pressure level (SPL)** is the logarithmic measure of the effective sound pressure of a sound relative to a reference value. It is defined as follows:*

$$L_p = 10log_{20}\frac{p}{p_0}dB$$

where p is the sound pressure, and p_0 is the reference sound pressure (usually taken to be 20

μPa, commonly referred to as the threshold of hearing).

Sounds intensity level (SIL) *is the level of a sound intensity relative to a reference value. It is defined as follows:*

$$L_p = 10log_{10}\frac{I}{I_0}dB$$

where I is the sound intensity, and I0 is the reference sound intensity, commonly taken to be $1pW/m^2$ in air at ambient temperature, where the acoustic impedance is $z_0 = 410Pa \cdot s/m$.

One can therefore speak of ***sound level*** *more generally provided that reference impedance and threshold of hearing values are specified. This holds since $I_0 = p_0^2/z_0$, and equating SIL and SPL requires that:*

$$\frac{I}{I_0} = \frac{p^2}{p_0^2}$$

Loudness *is the subjective perceptual psychoacoustical quantity that is correlated with the level, frequency content, and duration of a sound. Stated differently, loudness is "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud".*

Loudness level (in phons) *is the level of a sound judged to be equivalent to the level of a 1000 Hz reference tone (taken to be the average as assessed by healthy young adults with normal hearing). For example, a vowel sound with a loudness level of 40* ***phon*** *is judged to be equivalently loud as a 40 dB 1000 Hz sustained pure tone. By performing this across the range of perceivable frequencies at different loudness levels,* ***equal loudness curves*** *can be established (see Fig. A.3). A doubling in the number of phons does not generally result in a doubling of loudness. It is for this purpose that we define the* ***sone****. One sone is the loudness of a sound for which the loudness level is 40 phons, and the loudness of a sound that is judged to be n times that of a 1-sone tone is n sones (i.e. the scaling is linear). The relationship between sones and phons is described as follows:*

$$Phon = 40 + 10log_2(Sone)$$

As important as a solid foundational understanding of the acoustic signal and its related measures are, they are introduced here only for sake of their specific utility in helping to describe what is more germane to my thesis – namely, the speech signal. Before characterizing the quantities and qualities of speech that are relevant to my research, I will first introduce the source-filter model of speech production, as the latter is extremely helpful in

the understanding of the former.

A.2 The source-filter model of speech production

The source-filter model of speech production, illustrated schematically in Fig. A.4., is a good, simple, and widely used linear model of speech production. According to the model, the speech signal is synthesized by the convolution of a sound source (the vocal folds reacting to the airflow from the lungs) and a linear acoustic filter (the vocal tract – which is comprised of the oral cavity, tongue, lips and jaw). The source and filter components are assumed to be independent, and the subcomponents of the vocal tract are all assumed to be linearly additive. Neither of these assumptions holds true in general; however the model is nonetheless quite useful from a pedagogical perspective, and is a good first approximation to ground theories involving an understanding of auditory neuroscience. In non-tonal languages such as English, the source characteristics typically convey non-linguistic information, such as sex, emotional state, attitude, and speaker identity, whereas the filter characteristics convey the linguistic content or the 'message', and are experienced as the vowels and consonants that make up the code of language (Kraus & Nicol, 2005).

A.3 The speech signal - that's what I'm talking about!

The typical speech utterance is the result of a string of primitive acoustic patterns – most frequently simply vowels and consonants – produced by varying the configurations of the source and the filter to yield the targeted sounds. These basic units, known as phonemes, when taken altogether, constitute the building blocks of a language. These phonemes can be broadly categorized with reference to the nature of the source during production: namely, whether the source is periodic, noisy, or impulsive, or some combination thereof (Holt & Lotto, 2010; Quatieri, 2006). The phonemes typical of American English can be seen in Appendix F and are arranged in a tree in order to a provide a more nuanced understanding of their categorization. The stimuli of concern within the scope of my research all involve quasi-periodic sources, as they are all English vowels. Since English is not a tonal language, the space of English vowels can be partitioned with reference to their respective filter functions, which are dependent upon the shape of the vocal tract (Harrington & Cassidy, 2012). This very phenomenon is clearly illustrated in Fig. A.5 below. The first column of the figure shows different articulatory configurations as well as the phoneme that the configuration yields while acting on a voiced source; the second column shows a time domain waveform of the speech signal, which typically exhibit some quasi-periodic behaviour vis-à-vis their

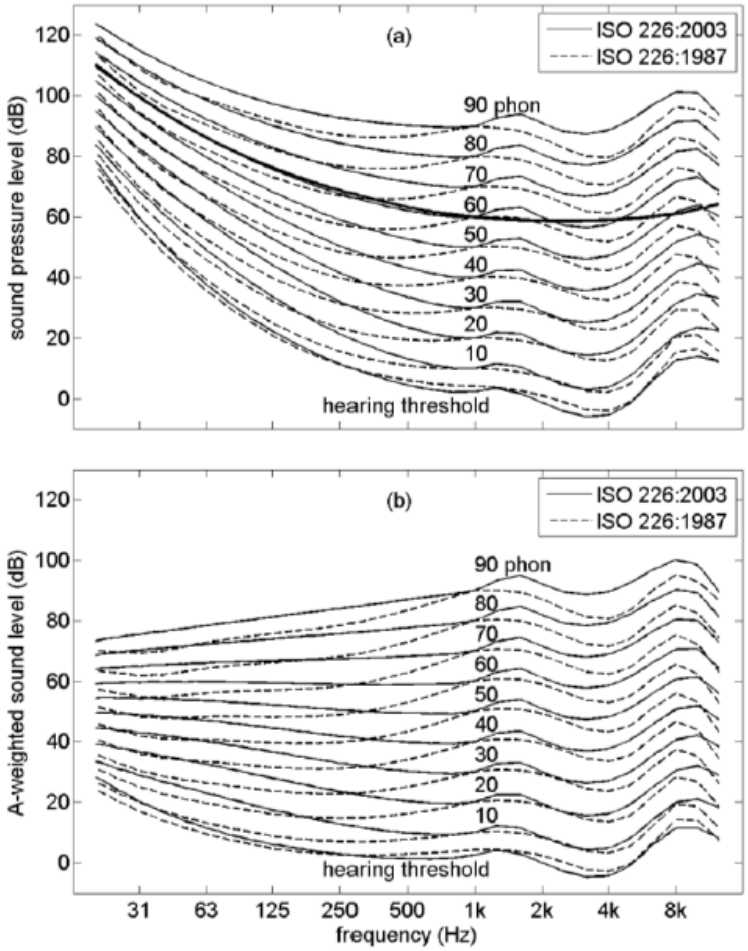

Figure A.3: Equal loudness curves. The curves represent equal loudness as perceived by the average human as a function of frequency and sound pressure level (a), and A-weighted sound level (b), and show the dependence of loudness assessments on frequency. The thick line represents the A-weighting curve. As an example, the curves labelled as 40 and 80 (phon) represent the sound level at which other frequencies are judged to be equivalent to 40 dB and 80 dB SPL of a 1000 Hz reference tone, respectively. Notice that since the ear is less sensitive to lower frequencies, a much greater SPL is required to elicit the same loudness as those frequencies nearer to the conversational speech range. This fact has driven the creation of A-weighting as a means of accounting for this frequency dependence on relative loudness. Reprinted with permission from "Practical Ranges of Loudness Levels of Various Types of Environmental Noise, Including Traffic Noise, Aircraft Noise, and Industrial Noise", by E. Salomons and S. Janssen, 2011, International Journal of Environmental Research and Public Health, 8, p. 1851. Copyright by E. Salomons and S. Janssen.

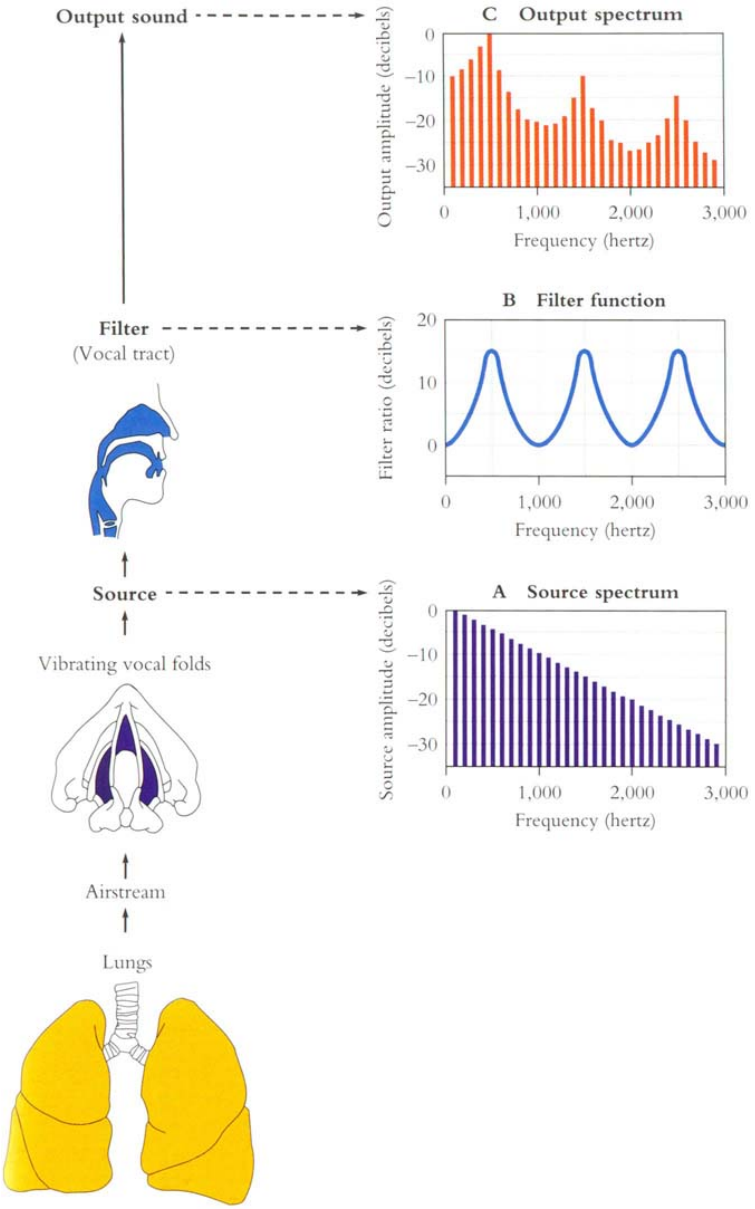

Figure A.4: The source-filter model of speech production. The air from the lungs causes the vocal folds to vibrate (source), and the vocal tract shapes the sound (filter) in accordance with the speech target. The frequency spectrum representation of both the source (A) and filter (B) and their convolution is shown (C). Reprinted with permission from Acoustic Phonetics, by V. Gramley, 2010, Retrieved from http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/acoustic-2010.html. Copyright 2017 Bielefeld University.

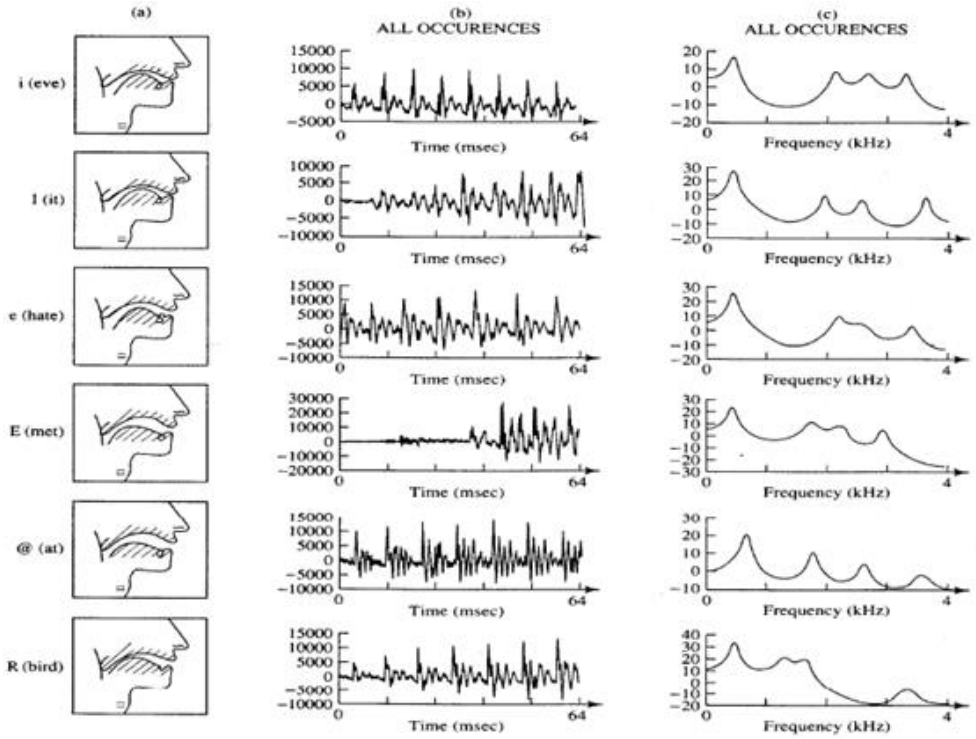

Figure A.5: English vowel production. A: Schematic representations of the vocal-tract (filter) profiles. B: The typical corresponding acoustic waveforms (i.e. time-domain representation that result. C: The corresponding amplitude spectrum (i.e. frequency-domain representation) for each vowel. Notice the characteristic formants and their relative unique spacing in frequency space. The first two or three formants are sufficient for accurate categorization. Reprinted with permission from Discrete-Time Processing of Speech Signals (p. 122), by J. Deller et al., 2000, New York, NY: John Wiley & Sons Ltd. Copyright 2000 IEEE.

amplitude deflections in time; and the third column shows the spectral content of the very same waveform.

The characteristic spectral peaks of these vowel phonemes – known as formants – are what enable our auditory system to differentiate between vowels (Harrington & Cassidy, 2012; Holt & Lotto, 2010). Psychoacoustical speech studies have revealed that two formant approximations to natural vowels – which may possess four or more distinguishable formants – are often sufficient for accurate identification and comprehension (Harrington & Cassidy, 2012; K. Johnson, 2011). Figure A.6 shows a number of American English vowels plotted along the typical frequency range of their first two formants (labelled on the two axes as F1 and F2, respectively). As can be seen from the figure, each vowel roughly occupies its own distinct area in this two dimensional formant space, and the compactness of the space hints at an approximate topological covering, which seems to indicate that the set of English vowels that have emerged make use of the frequency space that is most robustly encoded via our auditory system in order to maximize the accurate representation of the speech signal in light of the wide range of natural vocal variability between humans, while simultaneously minimizing the possibility of erroneous judgments of vowel identity.

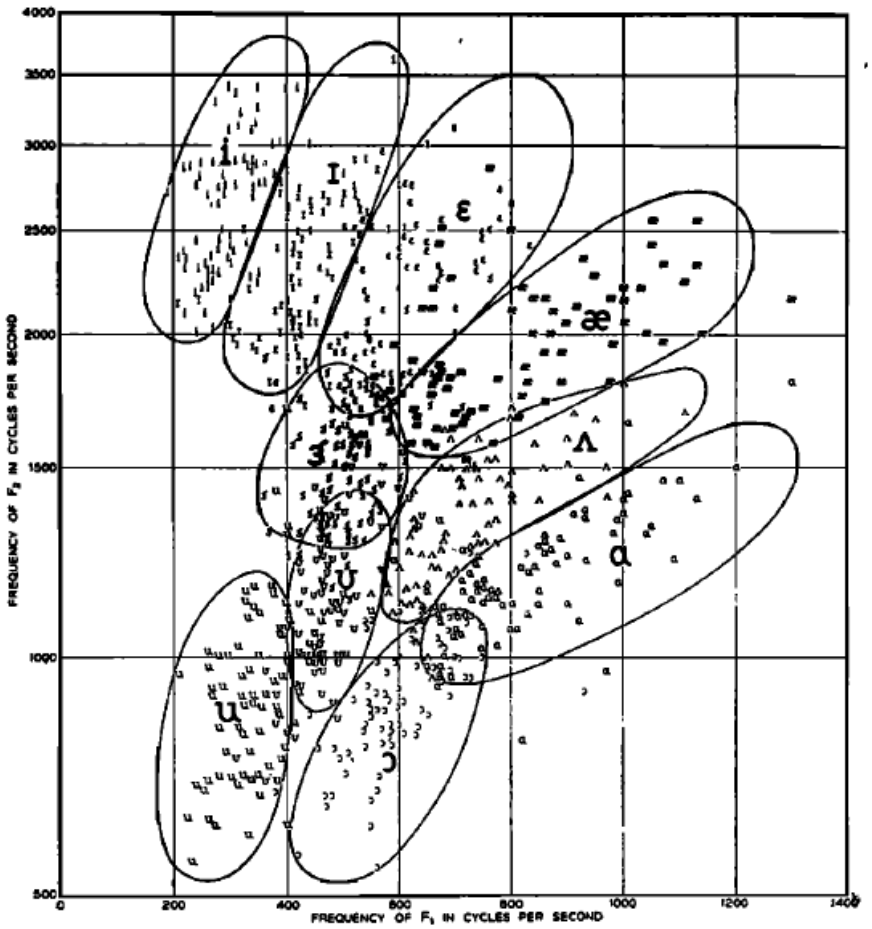

Figure A.6: Two-dimensional formant space, depicting the second formant (F2) versus the first (F1) for 76 normal subjects. Notice that the vowels partition the plane into a relatively compact, and minimally overlapping space, therefore allowing for accurate identification of each vowel based on these two formant features alone. Reprinted with permission from "Control Methods Used in a Study of Vowels", by G. Peterson and H. Barney, 1952, The Journal of the Acoustical Society of America, 24, p. 182. Copyright 1952 by Acoustical Society of America.

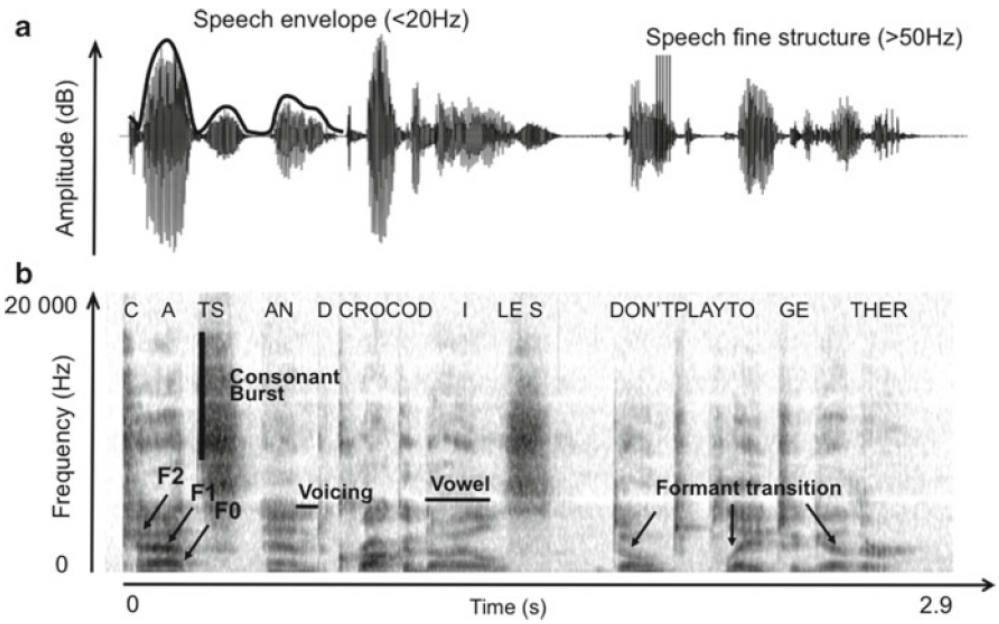

Figure A.7: The speech signal. (A) waveform and (B) spectrogram of an English sentence uttered by a male speaker. The speech envelope reflects the consistent, slow modulations that denote the syllabic structure, and appear to be sampled by the auditory cortex for parsing purposes. The speech signal has a second timescale concurrently at play, with modulation rates higher than 50 Hz representing fine spectral changes that signal speech sound identity and other relevant speech attributes. It is worth pointing out that the timescales indicated here are relative to the timescale under consideration, and in the case of a single vowel of consonant vowel pair, as typically used in frequency following response studies (see Sections 2.2.1 and 2.2.4 above), the envelope is usually taken to be from 50-250 Hz, and the temporal fine structure is often from 300-1200Hz. Reprinted with permission from "Speech Perception from a Neurophysiological Perspective", by A.-L. Giraud and D. Poeppel. In The Human Auditory Cortex (p. 228), by D. Poeppel et al. (Eds.), 2012, New York, NY: Springer New York. Copyright 2012 by Springer Science+Business Media LLC.

The act of perceiving speech can be viewed as a subroutine of the broader task of language comprehension, and it therefore requires no computation of meaning and can thus be approached by investigating the perception of speech sounds in isolation (e.g. at the phonemic or syllabic level, such as /a/ or /da/) (Giraud & Poeppel, 2012). As a result, many current models of speech perception are derived from studies of the perception of such single speech sounds (Holt & Lotto, 2010), and these same sounds therefore make for ideal candidates when investigating the neural correlates of the task of perception. Of course, it is prudent to bear in mind that natural language yields speech signals that are much more dynamic and full of contextual cues that are often not present in these basic speech units. In reality, the auditory system makes use of some subset of all of the available cues in order to index the sex and age of the speaker, emotional cues, regional and cultural characteristics, etc. in order to glean an exceptional amount of additional information beyond the mere computation of the literal meaning of the utterances themselves (Beigi, 2011). Figure A.7 provides a good example of a natural English speech signal for a sentence, and indicates a variety of features in both the time and frequency domain that are analyzed in order to accomplish this task.

A.4 The hardware and software of hearing – an overview of the anatomy and neural coding strategies of the auditory system

I now turn to a brief treatment of the anatomy of the auditory system (Fig. A.8 with a particular focus on the neural 'hardware' believed to be involved in the generation of the FFR and to what we know of the way that speech and speech level is encoded at each stage, as this will serve to inform the analysis of my own results.

A.4.1 The ear

The human ear is most typically divided up into three parts: the outer ear, the middle ear, and the inner ear. The outer ear consists of the visible, cartilaginous folds– known as the pinna – which function by attenuating and reflecting sound waves, thereby producing valuable cues related to elevation and direction of the sound source(s), as well as the auditory or ear canal through which sound waves enter the auditory system. The middle ear consists of the ear drum – formally referred to as the tympanic membrane – as well as three ossicles dubbed the malleus, the incus, and the stapes, or the hammer, the anvil, and the stirrup, respectively as a result of the similarity of their shape to these objects. As the sound waves propagate through the ear canal and finally reach the eardrum, the local changes in the medium's pressure cause the eardrum to vibrate, and these vibrations then drive the ossicles mechanically, ultimately inducing vibrations in the fluid of the inner ear as the footplate of the stapes correspondingly puts pressure on the oval window. The resonance properties of the outer and middle ear work to selectively and significantly amplify the acoustic signal, particularly around the 3 kHz range, as can be seen in Fig. A.9. This enhanced sensitivity helps to explain why humans are particularly susceptible to hearing loss in this very same range (Purves et al., 2012).

The inner ear is comprised of the snail-like structure known as the cochlea, as well as additional structures related to the vestibular system which is primarily responsible for providing information related to balance functions, and are therefore not directly related to hearing. Just as the middle ear is responsible for the transduction of acoustic energy into mechanical energy, the cochlea is the body responsible for the transduction of mechanical energy into electrochemical energy. When the aforementioned oval window is acted upon by the stapes, the deviations in its membrane induce movement of the fluid of the inner ear known as the perilymph, and begin the propagation of a travelling wave down the length of the basilar membrane (BM).

The base of the basilar membrane is narrower and stiffer than the apex, which results

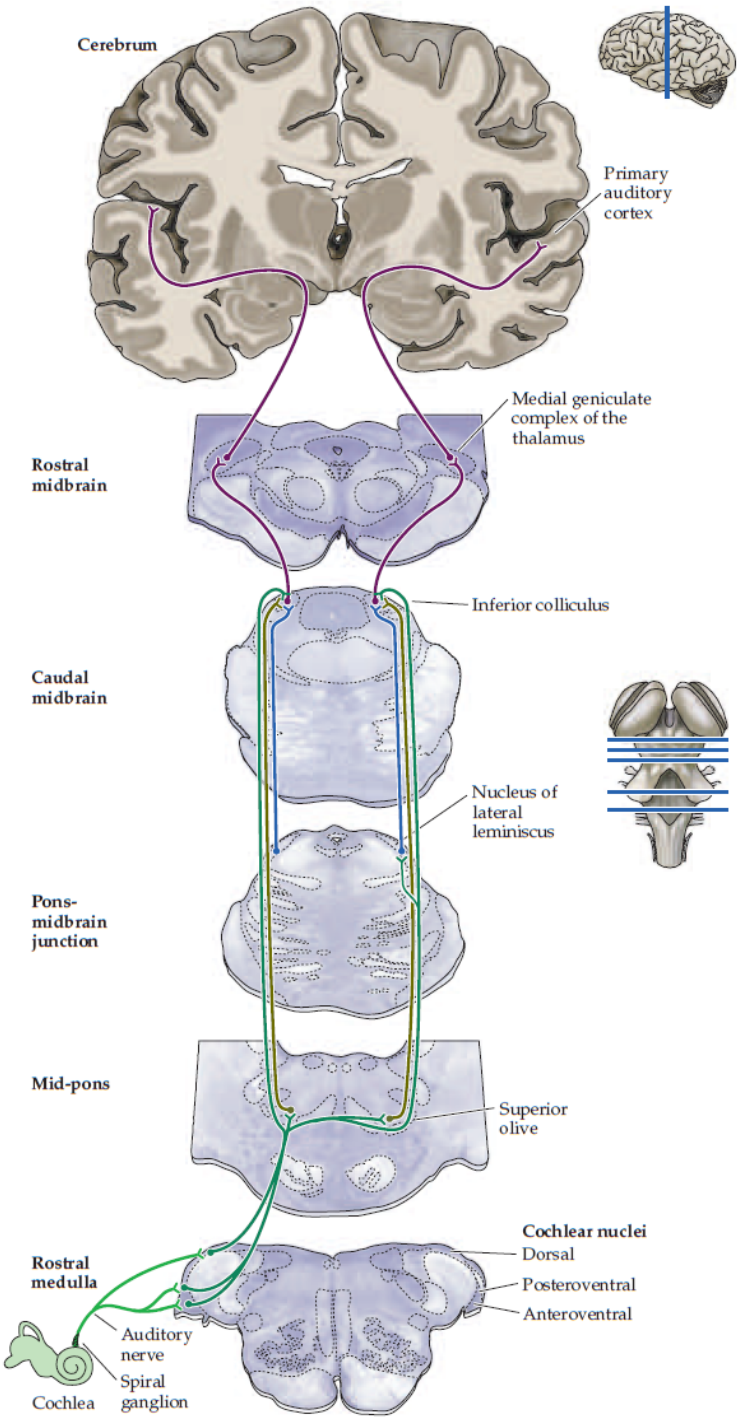

Figure A.8: The auditory system. This schematic depiction shows the principal central connections involved in hearing from the inner ear to the primary auditory cortex. Reprinted with permission from Neuroscience - 4th Ed. (p. 277), by D. Purves et al., 2012, Sunderland, MA: Sinauer Associates. Copyright 2012 by Sinauer Associates.

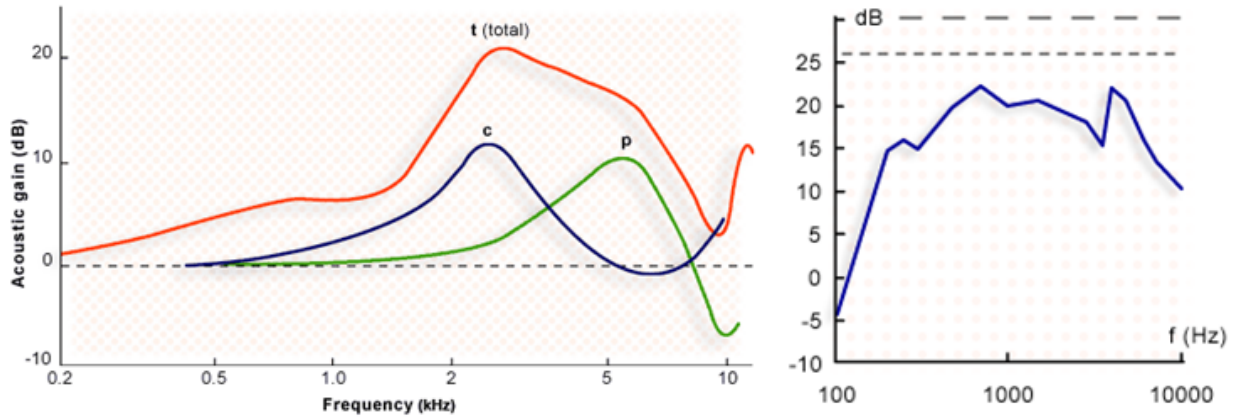

Figure A.9: Acoustic amplification of the outer and middle ear. A – The total (t) acoustic gain (dB) as a function of frequency (kHz) that results from the summation of the individual amplification profiles of the ear canal (c) and the pinna (p). B – The middle ear transfer function, displaying the gain (dB) as a function of frequency (Hz) that results from the mechanical action of the ossicles as they transduce the acoustic waveform into vibrations in the perilymph (fluid) in the inner ear. The combination of these two transfer functions explains the shape of the audibility curve shown in Fig. 29 (a) above. Reprinted with permission from Journey Into the World of Hearing, by P. Minary, 2016, Retrieved from http://www.cochlea.eu/en/. Copyright 2016 Association NeuroOreille.

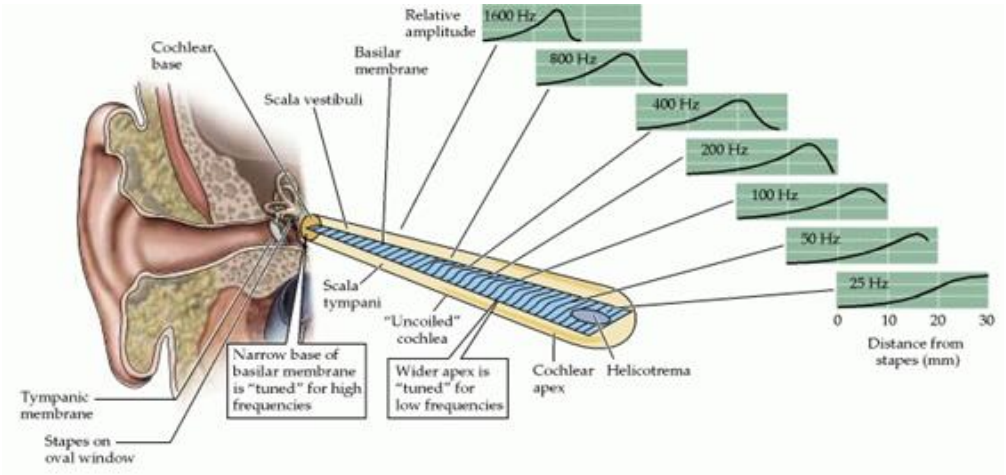

Figure A.10: The ear. Traveling waves along the cochlea (here uncoiled for pedagogical clarity). Notice the tonotopic organization, as the basilar membrane responds maximally to higher frequency sounds towards its narrow base, and to lower frequency sounds towards its wider apex. Reprinted with permission from Neuroscience - 4th Ed. (p. 277), by D. Purves et al., 2012, Sunderland, MA: Sinauer Associates. Copyright 2012 by Sinauer Associates.

in the basal end being resonant to higher frequencies, and the apex being resonant to lower frequencies, as is clearly illustrated in Fig. A.10 above. These changes in width and stiffness are gradual, and therefore produce a gradient in frequency from high to low along the length of the basilar membrane – a characteristic referred to as tonotopy that is maintained throughout much of the auditory system. It is this tonotopic organization that lends much support to the well-known place coding strategies that have been proposed, wherein the perception of sounds is attributed to the position of activation along the tonotopic axis (see Section A.6 below). This has led to the view that the inner ear behaves much like a Fourier analyzer, at least to a first approximation, by decomposing complex sounds such as speech into their constituent frequency components, leading to the construction of a variety of linear cochlear filterbank models over the years. A schematic overview of one such filterbank is shown in Fig. A.11. Despite the undoubted pedagogical utility of such linear models, it was known to many of the pioneering researchers investigating cochlear mechanics, such as von Bekesy, that the assumption of linearity does not hold (Olson, Duifhuis, & Steele, 2012; Von Békésy, 1963). In order to gain a better understanding of the nonlinear behaviour of the system, an understanding of the roles of the inner (passive) and outer (active) hair cells is necessary.

The IHCs – of which there are roughly 3500 – are the cellular units that are effectively responsible for the transduction of the mechanical energy that is propagated along the length of the BM into the electrical pulses referred to as action potentials that ultimately propagate or fire along the AN (see Section A.7 below). The travelling wave causes an up and down motion of the basilar and tectorial membranes at the same frequency as the stimulus, which results in a transverse shearing force that causes the deflection of the hair cell stereocillia in the vicinity of the force production, which thereby maintains the tonotopic organization of the system at the level of the AN. These shearing forces result in the preferential gating of neurotransmitters in one direction as a result of the mechanical properties of the IHCs, thereby resulting in what is effectively a half-wave rectification of the input.

All of the aforementioned cochlear processes are not sufficiently able to account for the observed sharpness of the frequency tuning in a healthy hearing human subject. It can therefore be reasoned that there must be an active mechanism at play that increases the sharpness of frequency tuning as observed, the cause of which is typically attributed to the function of the outer hair cells. Although the exact mechanism is not fully understood, it is believed that the OHCs change in cell stiffness, motility (Deo & Grosh, 2004; Ulfendahl & Flock, 1998), and in the length of the stereocilia (Hakizimana, Brownell, Jacob, & Fridberger, 2012) in order to selectively amplify the BMs response to sound. These mechanisms act as a form of positive feedback in the auditory system, as the activity of the OHCs effectively

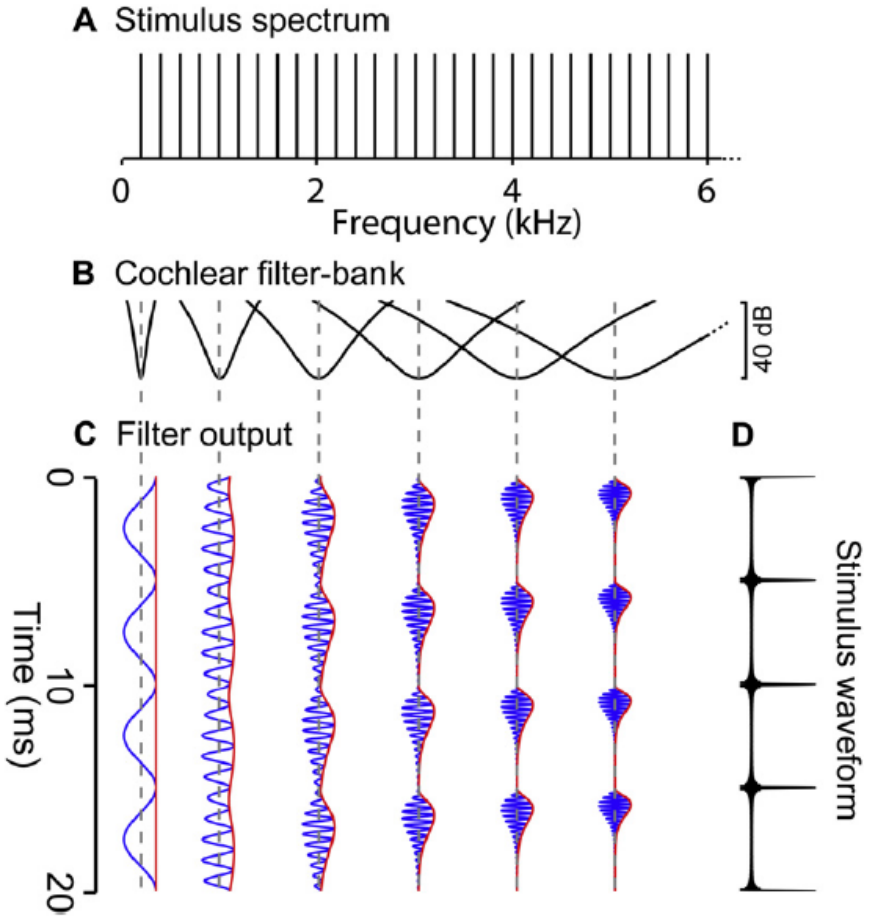

Figure A.11: The cochlear filterbank. A –a harmonic complex with F0 = 200 Hz. B – the cochlear filter-bank with centre frequencies of 0.2, 1, 2, 3, 4, and 5 kHz. Notice the increasing broadness of the filters with increasing frequency. This is derivative of the fact that the basilar membrane is better modelled as having logarithmic than linear in frequency selectivity. C – The filtered output in response to the stimulus in A. The blue lines represent the temporal fine structure of the stimulus, and the red lines represent the temporal envelope that is explicitly coded at the next stage of processing (see Figs. 42 and 43). With increasing centre frequencies, a corresponding increase in the temporal fine structure oscillations can be seen, and the temporal envelope becomes increasingly modulated as well. D – The stimulus waveform of the 200 Hz F0 complex, laid out to show the obvious fidelity of the encoding of its contents at the output of the cochlea and prior to its encoding in the ganglionic cells of the auditory nerve. Reprinted with permission from "Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds", by M. Sayles and I. Winter, Neuron, 58, p. 790. Copyright 2008 by Elsevier Inc.

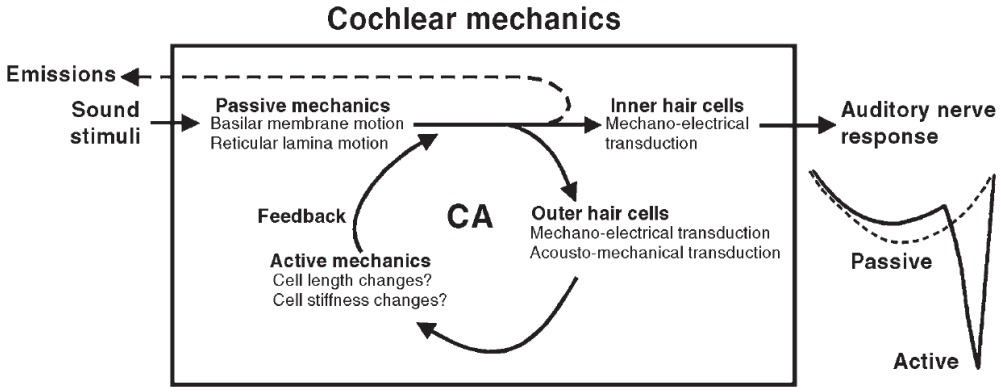

Figure A.12: The cochlear amplifier. A schematic depiction of the suggested role of the outer hair cell active (OHC) processes (cochlear amplifier, CA) in cochlear mechanics. A traveling wave produces passive motion of cochlear structures like the basilar membrane and the reticular lamina. The mechanical interaction with OHCs elicits changes in cell length, stiffness (i.e. electromotility) or stereocillia length, creating a feedback on the passive cochlear mechanics and enhancing the mechanical response of the system. This enhancement is detected by the inner hair cells which are in synaptic contact with afferent nerve fibres. The activity of the OHCs can also evoke emissions of sound into the ear canal, known as otoacoustic emissions. In absence of the CA, the neural response is broadly tuned (dashed curve), while its presence results in a response that exhibits both much greater sensitivity and selectivity (solid curve). Reprinted with permission from "Outer Hair Cells Provide Active Tuning in the Organ of Corti", by M. Ulfendahl and A. Flock, 1998, News in Physiological Sciences, 13, p. 111. Copyright 1998 by International Union of Physiological Sciences / American Physiological Society.

results in a sharpening of the tuning curve of the BM and the associated AN fibres as a result (see Fig. A.12 above). Importantly, the changes to the shape of the tuning curves of the BM are themselves level dependent, as is the compression ratio of the cochlea, which increases with increasing sound level, and underlies the emergence of some of the nonlinear distortion products typical of the auditory system (Gorga et al., 2007) (see Fig. A.13 below). Altogether, the behaviour of the cochlea is highly nonlinear and level-dependent, and the resulting neural representation of sounds is too as a consequence.

A.4.2 The neural coding space

All of the post-cochlear processing of speech occurs somewhere in the neural coding space, and so it is worthwhile to explore the possible ways in which the information available within the speech signal gets encoded via the transmission of neural impulses. Cariani (1999) provides a pedagogically useful schematic to outline the basic neural codes that define this space, as illustrated in Fig. A.14 below. These include channel codes (e.g. the aforementioned place code), temporal codes, or some hybrid thereof. Although we do know that tonotopy is maintained throughout the majority of the auditory system, and likely up to the level where the signal is transformed into features that are fed into the neural networks responsible for informational functions such as classification, there is strong evidence to suggest that a purely rate-place channel coding scheme is unlikely, as the representation of vowel formants in the auditory nerve's firing rate contrasts is no longer apparent at higher sound levels typical

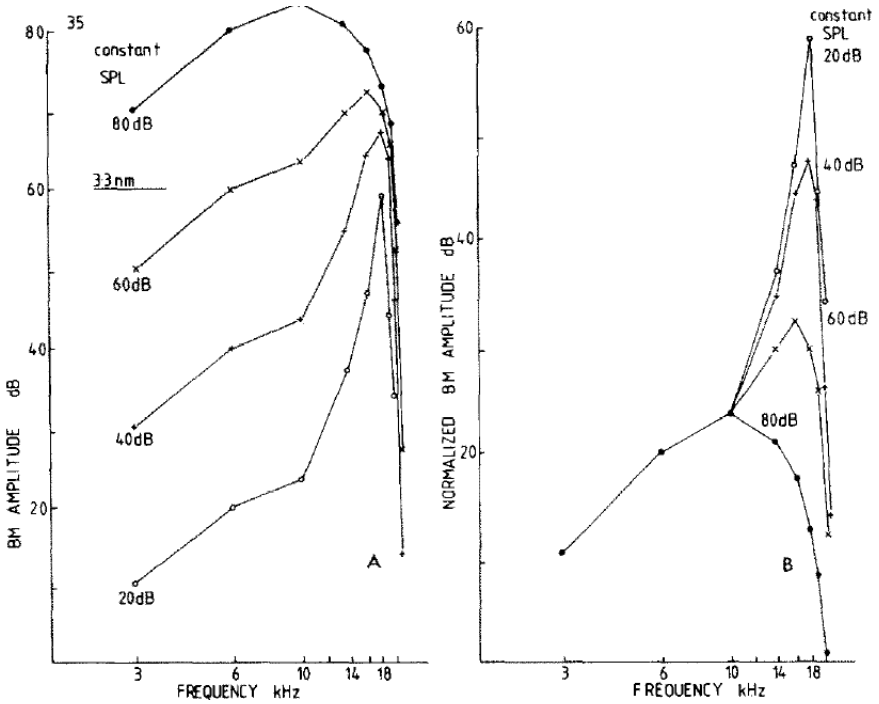

Figure A.13: Derived BM response envelopes at four sound pressures. A – The amplitude of the BM in dB is plotted versus frequency at four different sound levels (20, 40, 60, and 80 dB SPL). Notice that the frequency of maximum response shifts basally with increasing sound level, approximately corresponding to a half octave drop from the 20 dB to 80 dB presentation levels for a constant frequency. B – The same as A, but with each curve shifted by an amount equal to the difference between the BM amplitude and the presentation level. The observable gain compression is the result of a clearly nonlinear system, as the curves would be superimposed in a linear case. These compressive effects can be observed, for example, by noticing that the 80 dB SPL curve has a maximum normalized BM amplitude of roughly 25 dB, whereas the 20 dB SPL curve has a maximum normalized BM amplitude of roughly 60 dB. Reprinted with permission from "Basilar membrane measurements and the travelling wave", by B. Johnstone et al., 1986, Hearing Research, 22, p. 150. Copyright 1986 by Elsevier Science Publishers B.V.

of the speech range. It is worth cautioning that this might be accounted for by the high spontaneous rate nerve fibres, and so a strict place-based encoding cannot be discounted entirely; however, there is ample evidence to suggest that a hybrid, spatiotemporal pattern code is utilized in the encoding of speech, as shall be discussed in Section A.7. It is also very likely that several different coding schemes are employed throughout the auditory system to serve a variety of different purposes. For example, studies suggest that latency-place coding is utilized when analyzing sounds across ears in order to perform source localization tasks, as the ear nearest to the sound source will receive the sound before the more distal ear, and therefore this information can easily be exploited by a very temporally sensitive coincidence detector style of neural network (Franken, Roberts, Wei, Golding, & Joris, 2015).

A.4.3 The auditory nerve

The auditory nerve (often referred to as the cochlear nerve) is part of the eighth cranial nerve, known as the vestibulocochlear nerve. The AN, which consists of roughly 30,000 nerve fibres

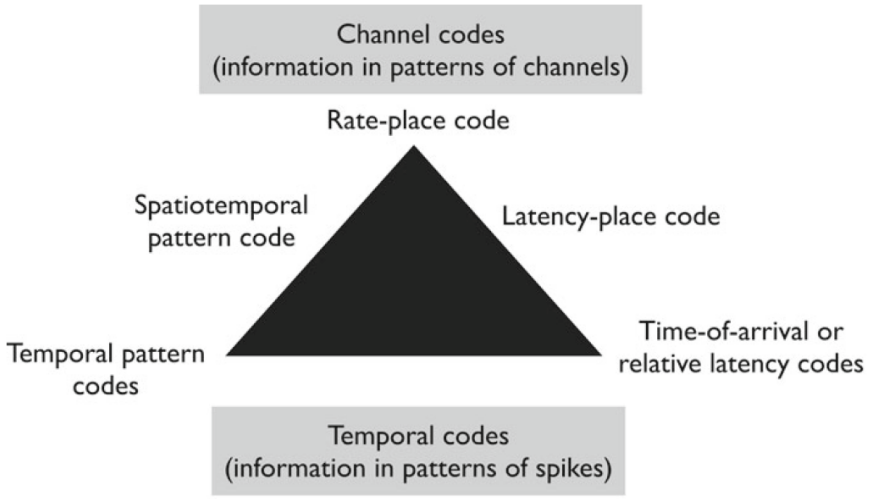

Figure A.14: Neural coding. A schematic depicting the space of basic neural pulse codes, which are divided into channel codes (i.e. place codes), temporal pattern code, and relative latency codes, although any combination thereof is feasible, as indicated along the edges. Reprinted with permission from "Toward a Theory of Information Processing in Auditory Cortex", by P. Cariani and C. Micheyl. In The Human Auditory Cortex (p. 352), by D. Poeppel et al. (Eds.), 2012, New York, NY: Springer New York. Copyright 2012 by Springer Science+Business Media LLC.

in normal human subjects, is the pathway responsible for the neural transmission of auditory sensory information to the rest of the brain. The vast majority of these fibres – some 90-95% – are fibres that innervate the IHCs of the cochlea; the remaining 5-10% innervate the outer hair cells. The cell bodies of the cochlear nerve lie within the cochlea proper, with their peripheral axons connecting with hair cells via ribbon synapses, which promote rapid neurotransmitter release and sustained signal transmission. AN fibers discharge as a result of the depolarization of the hair cells, which in turn cause the diffusion of neurotransmitters across the junction between the hair cell and the nerve fibre, where they bind with receptors on the fibre and cause a depolarization that discharges an action potential down the length of the peripheral axon, across the soma of the fibre, and finally down along the central axon of the unit. These central axons of the AN fibres pass through the internal auditory canal and ultimately terminate as they synapse ipsilaterally with cell bodies of the cochlear nucleus – the most caudal aspect of the auditory brainstem.

Figure A.15 highlights the bimodal distribution of the AN fibres, which can be divided up into high and low-medium spontaneous discharge rates (NB: classically they have been partitioned into three rate classes in the literature – high, medium, and low – but the statistics do not veritably reflect this trimodal conceptualization). The net effect of such a distribution is that the large dynamic range of naturally occurring, behaviourally relevant sounds that humans encounter can be fully encoded in the AN. The high spontaneous rate fibres allow for sensitivity in the detection of very soft sounds, but their utility in encoding sounds above 40 dB SPL is minimal, and so the task is believed to largely be undertaken

by the low-SR fibres, which are far fewer in number, and which are much more prone to cell death when faced with persistent noise, even at fairly low sound levels (Kobel, Le Prell, Liu, Hawks, & Bao, 2017). The death of these low-SR fibres has very recently become a topic of considerable interest, as the phenomenon is thought to be the cause of so-called hidden hearing loss, wherein subjects possessed of otherwise normal audiological function have degraded ability to perceive and comprehend speech in noise. This notion certainly has surface validity, as the high-SR fibres are necessary and sufficient for detecting very low level sounds, and so this hearing loss due to the low-SR fibres would indeed be hidden by an audiological screening composed of little more than an audiogram.

There exists a rich body of literature on the subject of speech encoding in the AN across a wide range of animal models (see Delgutte (1980) for a good introduction). The vast majority of such studies involve exposing single fibres with known CF's and filter profiles to the entire stimulus waveform, often across a variety of sound levels. Figure A.16 below is demonstrative of such studies, and it is very instructive for the purpose of understanding encoding vowel formant(s) and the envelope at the fundamental (F0) of the speech signal (see the caption for a description of these phenomena). Although both fibres show clear phase-locked activity to the F1 frequency at most levels, it is clear by inspection that the fibre with a CF much higher than F1 also robustly encodes the amplitude modulation at F0 when the stimulus level is sufficiently high. The differential effects of sound level on these encoding trends are also visually apparent, but it is worth pointing out that there appears to be a saturation occurring in terms of the absolute amplitude of the discharge rate at higher levels, and the depth of modulation clearly decreases for both fibres with increasing level. Since both are high-SR fibres, this is well aligned with the notion that the low-SR fibres are likely to play an important role in encoding of relevant features at higher levels – and particularly envelope features.

Studies that have sought to characterize the AN's response to vowel stimuli more broadly across frequency have also been carried out. Figure A.17 is highly informative, as it demonstrates how the first two formants of a synthetic vowel are robustly encoded centered around the fibres whose CFs are near to the formant frequency, whereas F0 is very clearly encoded at the lower frequencies, and re-emerges in the higher frequency band above those fibres that largely capture F2. This figure also highlights the complex phenomenon related to the relative lag of different frequency bands as the AN innervates the cochlea according to its tonotopic arrangement, which results in higher frequency regions responding to the stimulus earlier than the more apical lower frequency regions (see Fig. A.10). This is especially important to bear in mind when assessing electrophysiological results in the time-domain, such as those due to the FFR.

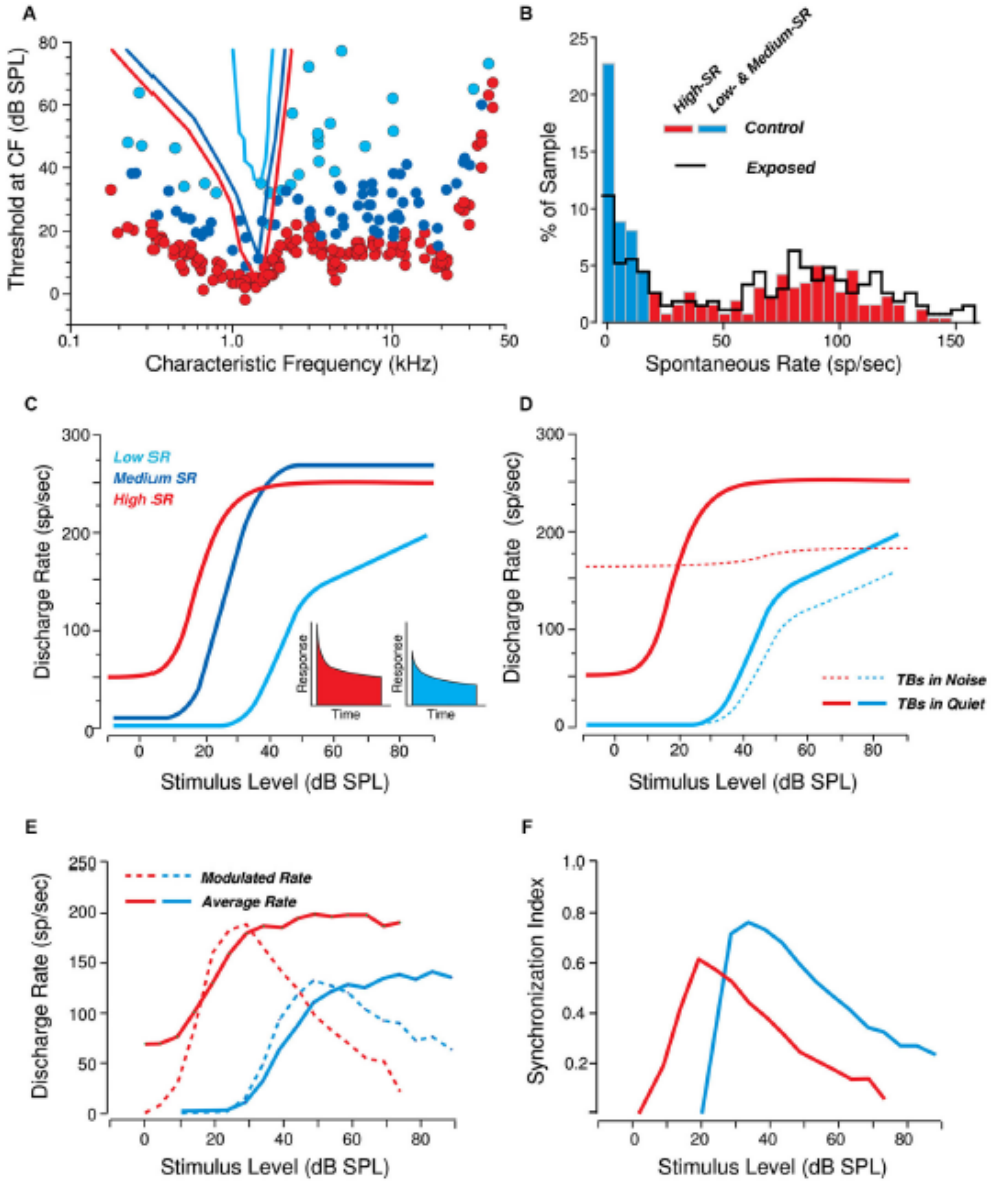

Figure A.15: The bimodal distribution of afferent auditory nerve fibres by discharge rate. A: threshold tuning curves of example high, medium, and low SR fibres (key shown in C) superimposed on a scatterplot of thresholds at the CF from a cat. B: Distribution of spontaneous rates in AN fibres in pig before and after (black line) noise exposure causing a reversible elevation of thresholds .C, D: Schematic rate-level functions to tone bursts at the CF in quiet (C), and in noise (D). E, F: Responses to SAM tones high vs. low fibres, expressed as average rate and modulated rate (E) or average synchrony (F) (carrier tones are at CF, AM of 100 Hz). Reprinted with permission from "Cochlear neuropathy and the coding of supra-threshold sound", by H. Bharadwaj et al., 2014, Frontiers in Systems Neuroscience, 8, p. 4. Copyright Bharadwaj et al. (2014) by Baradwaj, Verhulst, Shaheen, Liberman and Shinn-Cunningham.

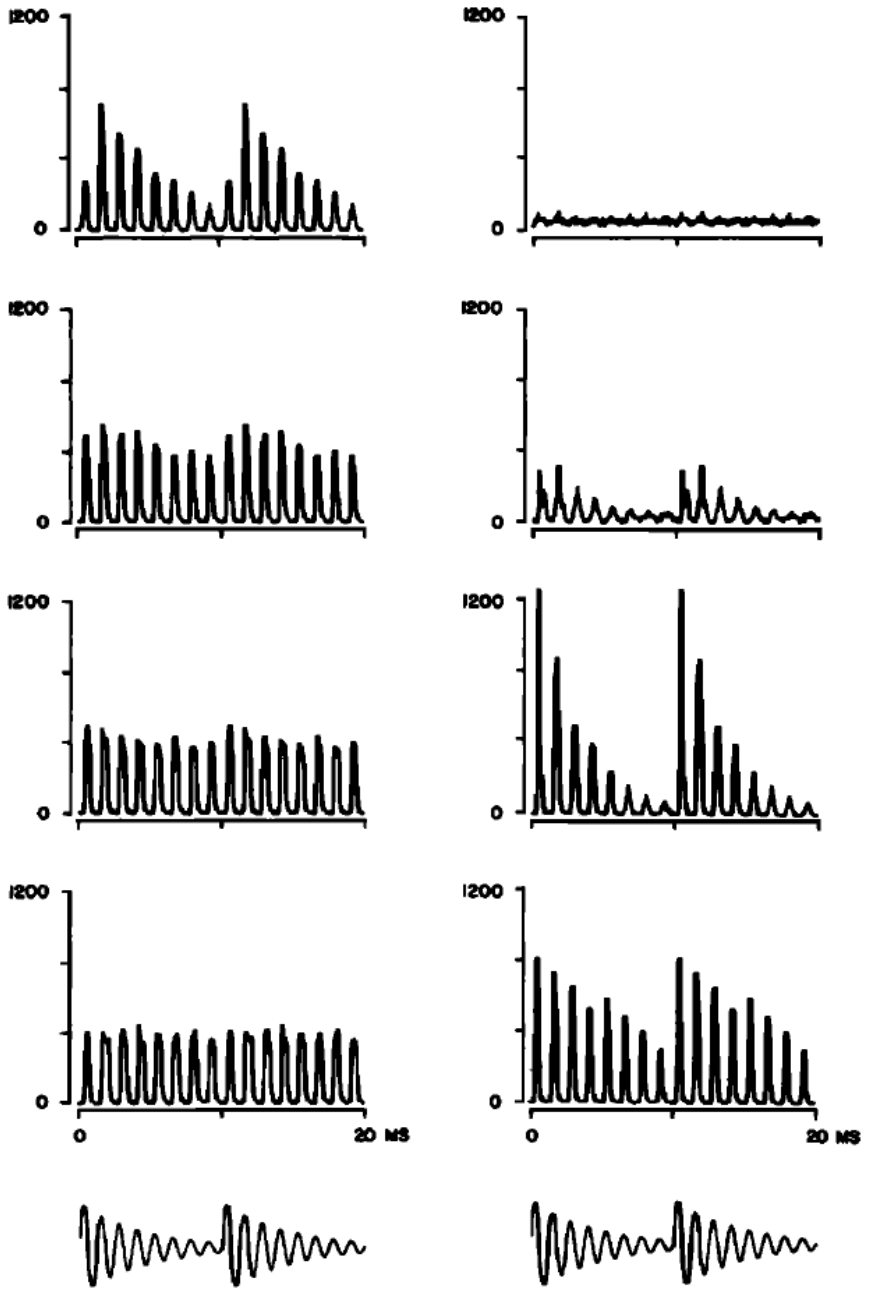

Figure A.16: Auditory nerve unit response to a single formant vowel. The response patterns of two auditory nerve fibres to a single-formant, synthetic vowel (F0=100 Hz, F1=800 Hz) as recorded over two periods of the acoustic stimulus (shown in the bottom panel) at levels of 32, 47, 62, and 77 dB SPL, as labelled. The fibre on the left has a CF of 780 Hz, which is very near to the formant, and the fibre on the right has a CF of 2790 Hz. The responses from each unit are presented as period histograms with 50 microsecond bin widths. Both units show strong discharges to the formant frequency, as the peaks of the response patterns are in all cases separated by 1/F1, but differ in their envelopes at all stimulus levels. The unit on the left exhibits clear modulation to the fundamental frequency at low levels, which diminish with increasing level until they flatten out or saturate at the highest level (consistent with the rate-level profile of AN fibres). The unit on the right exhibits strong fluctuations at the fundamental across all presentation levels, as its CF is sufficiently far from the main frequency components of the stimulus, and does therefore not reach saturation even at the highest level. This is characteristic of vowel coding across the auditory nerve. Reprinted with permission from "Representation of speech-like sounds in the discharge pattern of auditory-nerve fibers", by B. Delgutte, 1980, Journal of the Acoustical Society of America, 68, p. 848. Copyright 1980 by Acoustical Society of America.

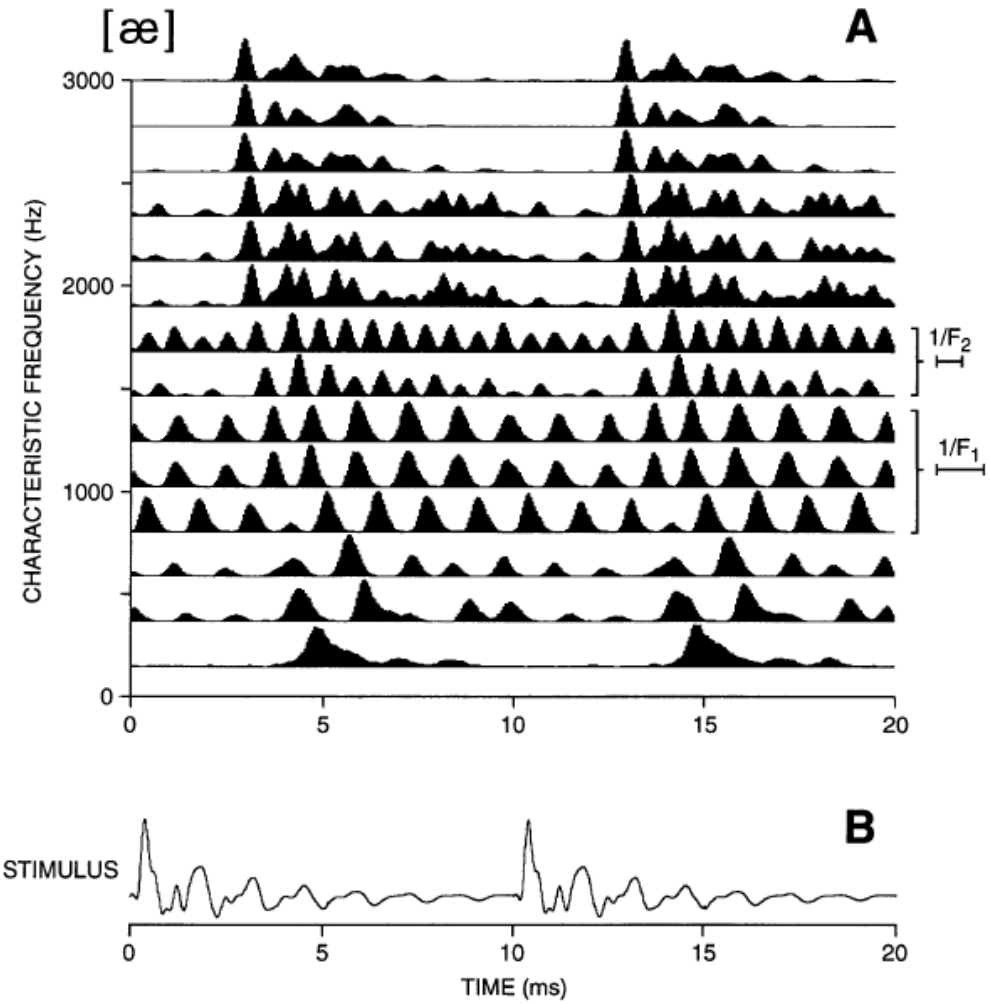

Figure A.17: The response to a synthetic two-formant /ae/ vowel across the auditory nerve. The vowel is shown here presented over two periods at a level of 60 dB SPL, with an F0=100 Hz. While fibres below F1 and above F2 are most strongly synchronized to F0, the fibres with CF around each formant lock to those specific formant frequencies, with some evident envelope modulation at F0. Note the relative latency between the stimulus onset and the response, and in particular the shorter latency in the response at the mid and higher frequency (basal) end of the cochlea when compared to the lower frequency region. Reprinted with permission from "Auditory Neural Processing of Speech", by B. Delgutte. In The Handbook of Phonetic Sciences (p. 520), by W. Hardcastle and J. Laver (Eds.), 1999, Oxford: Blackwell. Copyright 2010 by John Wiley and Sons Inc.

Most hearing impairments are the result of peripheral damage of some sort, either through damage to the inner or outer hair cells of the inner ear (i.e. sensorineural hearing loss (SNHL)), or through AN fibre loss due to prolonged noise exposure (i.e. hidden hearing loss (HHL)). Unsurprisingly, most hearing-aid design strategies aim to restore normal sound perception regardless of its genesis, since the end-user is really concerned with their ability to communicate and interact with their environment and much less so with the strategy required to get there. Although this seems like the right approach at first glance, an alternative strategy has been proposed: namely, the restoration of (more) normal AN firing patterns (Giguère, 1998; Giguère & Smoorenburg, 1999; Sachs et al., 2002).

The appeal here results from that fact that the AN is the sole conduit of ascending information to the rest of the auditory system, and so whatever information is available to these more rostral sites must be encoded somewhere in the AN's firing patterns. It therefore follows that if one is able to amplify the inputs of the hearing aid in such a way that the AN firing patterns are identical to what they would have been prior to insult, then all higher level perceptual and experiential qualities and related abilities would return to their pre-impairment baseline. Of course, such a complete restoration of normal function is very difficult to envision barring repair at the cellular level, and it is not even possible to amplify a pure tone in such a way that the AN firing pattern will behave normally in practice (Giguère, 1998; Giguère & Smoorenburg, 1999); however, the restoration of behaviour that approximates normal representation of key features is possible (Sachs et al., 2002). One must further disregard any concerns here related to neural plasticity that obviously hamper the surface validity of this hypothesis across the HI population. Barring strategies that actually make use of this plasticity in order to remap frequency space (e.g. frequency compression of transposition – see (Dillon, 2012, pp. 239)) and excluding more invasive technologies such as auditory brainstem implants that necessarily bypass the periphery, this strategy and its derivatives is certainly appealing and holds great promise in the quest for new HA signal processing strategies (Sachs et al., 2002).

A.4.4 The auditory brainstem

The auditory brainstem consists of a variety of specialized nuclei and fibre tracts that are dedicated to performing parallel computations of features relevant to various behavioural tasks from sound localisation to pitch and level coding, and likely the specific coding of complex sounds such as speech (Middlebrooks, 2014). Figure A.18 below depicts our current understanding of some of the parallel pathways present in the cochlear nucleus alone. Of greatest relevance to the FFR (and in ascending order) are the cochlear nucleus (CN), the

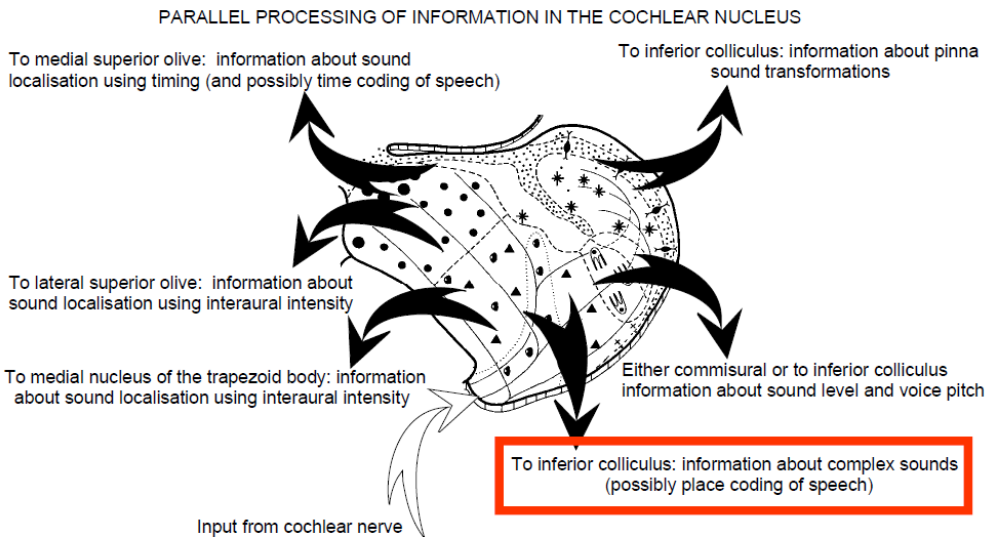

Figure A.18: Parallel processing in the CN. A schematic of the various parallel processing pathways of the cochlear nucleus. These include pathways devoted to sound localization, pitch and level processing, and speech processing (highlighted in red). Reprinted with permission from "Processing in the cochlear nucleus" [PowerPoint Slides], by A. Palmer (N.D.), Retrieved from https://health.uconn.edu/meds5377/wp-content/uploads/sites/151/2017/07/Palmer-Cochlear-nucleus-1-3.pdf. Copyright 2009 A. Palmer.

lateral lemniscus (LL), and the inferior colliculus (IC)(Skoe & Kraus, 2010). The cochlear nucleus consists of three tonotopically organized subcomponents: the ventral components known as the posteroventral cochlear nucleus (PVCN) and the anteroventral cochlear nucleus (AVCN), and the dorsal cochlear nucleus (DCN) (Middlebrooks, 2014). The patterns of synaptic convergence from multiple AN fibres onto CN targets is known to produce enhanced synchrony and to account for the dynamic range coding problem encountered in the inner ear (Joris, Carney, Smith, & Yin, 1994; Joris, Smith, & Yin, 1994). In mammals, there is also evidence that specialized neurons known as chopper units are capable of coding formant-to-trough spectral attributes across the range of levels typical of speech (Recio & Rhode, 2000), and that low-SR primary-like neurons in particular yield better representations than do the high-SR primary-like units at all but the lowest speech levels (May, Prell, & Sachs, 1998).

The lateral lemniscus receives a variety of both ipsilateral and contralateral inputs from the superior olivary complex (SOC) and the CN, and possesses three nuclei (Middlebrooks, 2014). The function of the LL nuclei are not generally known, although the LL is known to be very sensitive to changes in timing and amplitude of stimuli, and is thought to be involved in the computation of duration of complex sounds and possibly in the lateralization and localization of sounds owing to the convergence of binaural inputs (Middlebrooks, 2014). It does of course serve the major function of acting as a relay to the inferior colliculus for a large variety of converging inputs.

The IC itself is composed of four parts: the external and dorsal cortices, the dorsalmedial nucleus and the central nucleus (CNIC) (Middlebrooks, 2014). The cortices do not possess

the characteristic tonotopic organization of many of their more caudal structures (Middlebrooks, 2014), and is believed to be involved in the complex analysis of auditory information, such as speech processing (Syka, 2010). In fact, evidence from electrophysiological recordings of the IC in rabbits provides evidence that vowel formants are strongly encoded across a wide range of sound levels and in noise via neurons that are tuned to low-frequency amplitude fluctuations (Carney, Li, & McDonough, 2015). The central nucleus is tonotopically organized into laminae, and is believed to be involved in pitch and spectral processing, as well as in decoding interaural time and intensity differences for the purposes of sound localization (Middlebrooks, 2014).

Invariance of the neural representation of speech content to sound level changes is obviously extremely important for stable communication abilities, but some parallel mechanism must necessarily exist to encode and tag the level and loudness of these very same speech tokens. Simple firing rate codes of level require extensive transformations to become efficient since the ascending inputs alone cannot sufficiently generate a monotonically increasing rate over the whole ~110 dB range of normal hearing. It is unsurprising in light of this fact that there is evidence for a hierarchical approach to level coding across the mammalian brainstem and midbrain that consists of top-down and bottom-up interactions between a variety of driving and modulating components, and that highly non-linear responses to both noise and tonal stimuli have been observed throughout these areas (Grimsley, Sanchez, & Sivaramakrishnan, 2013).

Importantly, the IC is the first nucleus where neurons exhibit rate coding related to band-pass and band-reject filtering, whereby the firing maximum and minimum, respectively, are tuned to a modulation frequency, in addition to exhibiting the more temporally synchronized responses typical of caudal structures (e.g. CN) (Rabang et al., 2012). Thus, the IC is clearly involved in the sort of sophisticated feature extraction that is known to be required for the more high-level functions of the auditory system, such as categorization. This type of feature extraction can be seen in Fig. A.19 below, as comparisons between the encoding of speech tokens with a similar /da/ portion reveal enhanced contrasts between IC and CN neurons.

A.4.5 The thalamus and the cortex

The fibres from the IC form afferent connections to the thalamic relay known as the medial geniculate body (MGB), which has dorsal, medial, and ventral subdivisions (Rabang et al., 2012). The ventral region maintains the tonotopic organization of their CNIC inputs, and is believed to be involved in frequency analysis, as well as integrating intensity and binaural latency cues (Bartlett, 2013). The dorsal and medial subdivisions are believed to be involved

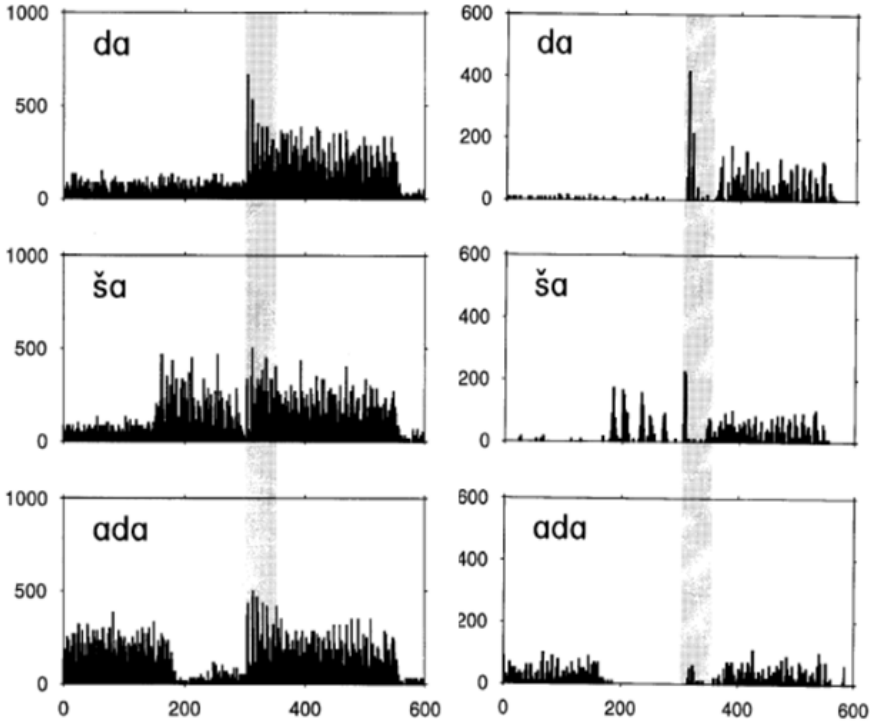

Figure A.19: CN and IC unit responses to speech. Peristimulus time histograms of single unit responses to three synthetic speech stimuli (presented at 60 dB SPL)in the CN (left) and IC (right). The stimuli are all spectrally aligned so that the /da/ segment occurs at the same time. This is true even of the /sa/ stimulus indicated in the middle panel, although it is not heard as /da/ due to the preceding spectral content thereby imparting a sense of the context-dependent encoding of speech. Notice the relatively sparse, yet increasingly complex encoding of the contrastive features in the IC response in comparison to the CN response to the same stimulus. In general, there is evidence for increasingly complex neuronal coding interactions at each successive stage in the ascending auditory pathway, clearly suggestive of inhibitory and facilitatory mechanisms that ultimately act as feature detectors, which create a basis for our categorical perceptual abilities. Reprinted with permission from "Auditory Neural Processing of Speech", by B. Delgutte. In The Handbook of Phonetic Sciences (p. 532), by W. Hardcastle and J. Laver (Eds.), 1999, Oxford: Blackwell. Copyright 2010 by John Wiley and Sons Inc.

in multimodal integration (Bartlett, 2013). Neurons in MGB are known to exhibit a variety of responses believed to subserve the processing of vocalizations, including onset, sustained, and offset responses (Bartlett, 2013), and a large proportion of its cells respond to individual types of vocalizations (Syka, 2010). Interestingly, a high percentage of MGB neurons have non-monotonic firing rate profiles with increasing sound level, which is hypothesized to contribute to the maintenance of sensitivity to a given sound level contrast in backgrounds of different sound levels, and may be critical for the perception of complex sounds such as speech (Bartlett, 2013). One study on the level functions of single unit responses in the MGB of cat suggests that sound level might be encoded by the mean firing rate of a population of monotonic units, or place coded by the distribution of maximally activated non-monotonic units which are broadly tuned to different levels (Rouiller, De Ribaupierre, Morel, & De Ribaupierre, 1983).

The auditory cortex consists of three major subdivisions: the core, the belt (which surrounds the core), and the parabelt (Rauschecker, 2015a). These subdivisions are functionally distinguishable via their neuronal behaviour as well, with the core showing narrow tuning to frequency, the belt responding best to band-passed noise of a specific bandwidth and frequency, and the parabelt responding to more complex sounds (Rauschecker, 2015a). The ventral stream of the belt appears to be tasked with auditory pattern and object recognition, which includes the decoding of speech sounds at the level of phonemes, words, and short phrases; the dorsal stream is involved with processing of auditory space and motion and is believed to be primarily concerned with sensorimotor integration and control, and is also believed to be involved in sentence comprehension (Rauschecker, 2015a).

The cortex is typically associated with the emergence of percepts as experienced by the subject (Frégnac & Bathellier, 2015), and there is evidence to suggest that the auditory cortex is no exception here. Despite a great deal of preprocessing by subcortical structures in humans, conscious auditory processing, including the perception, storage, and recognition of complex sounds such as speech and music, is thought to occur mostly at the level of the auditory cortex (Rauschecker, 2015b). This appears to hold true with respect to the emergence of the perception of the loudness as well (Behler & Uppenkamp, 2016a, 2016b; Röhl & Uppenkamp, 2012; Uppenkamp & Röhl, 2014). One recent fMRI study found that the percentage change from baseline neural activity at the level of IC within a subject when presented with an 80 dB SPL tone was not correlated with the subject's respective categorical loudness assessments, and that although a weak correlation seems to emerge at the level of MGB, it did not reach statistical significance. In contrast, a strong correlation was found at the cortical level, as demonstrated in Fig. A.20 below (Röhl & Uppenkamp, 2012). This supports the notion that level coding might be transformed into loudness coding

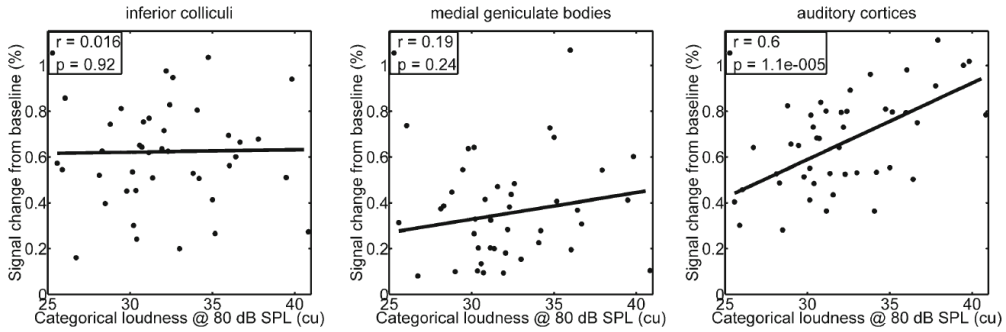

Figure A.20: The emergence of loudness. Here, the ordinate represents the percent signal change in the blood oxygen level-dependent effect relative to baseline, and the abscissa is represents the categorical loudness assessments, measured in categorical units (cu), of the subjects for a fixed presentation level of 80 dB SPL. This fMRI study suggests that there is no relationship between categorical loudness assessments and signal change from the baseline signal at the level of the inferior colliculus (r=0.016, p=0.92). Although a weak positive correlation was observed at the level of MGB, it is not statistically significant (r=0.0.19, p=0.24) and is relatively weak in comparison with that of the AC (r=0.6,p<0.001), suggesting that level coding is somehow translated to loudness coding as the neural response to a sound passes from IC to AC. Reprinted with permission from "Neural Coding of Sound Intensity and Loudness in the Human Auditory System", by M. Rohl and S. Uppenkamp, 2012, Journal of the Association for Research in Otolaryngology, 13, p. 375. Copyright 2012 by Association for Research in Otolaryngology.

between IC and AC. Put differently, there is no simple mapping between objective 'level' and the subjective experience of 'loudness'. There are likely to be a series of transformations and contextual elements that ultimately influence the perceived loudness of an auditory object. From what has been ascertained, the activity encoded in the IC is much more highly correlated with presentation level than it is with the perceived loudness, whereas the activity of the cortex is much more highly correlated with the perceived loudness than it is with the presentation level. At some point, what is initially encoding level must ultimately be rendered to the code for loudness. Other studies support the existence of level-invariant representations of sounds in the primary auditory cortex, which is a crucial component of robust feature recognition in natural environments (Sadagopan & Wang, 2008).

More generally, the cortex (beyond merely the auditory cortical areas) is believed to be responsible for the attentional modulation of sound objects and streams, and it is believed that the oscillatory sampling of the sensory systems (e.g. via theta, delta, or gamma rhythms) (VanRullen & Dubois, 2011) underlies the process of categorical perception (Gross et al., 2013). Recent efforts to simulate the types of cortical networks required in order to achieve such representations, wherein the behaviourally relevant information is sufficiently recovered and that this process is generalized across irrelevant stimulus variation (e.g. extracting the lexical information regardless of level), indicate the plausibility of the type of hierarchical processing scheme that appears to be employed in the auditory brain (Młynarski & McDermott, 2018; Tanaka, 2016). In addition to categorical perception, there are a variety of cognitive dimensions that involve cortical network processing downstream from the auditory cortex. These include mnemonics, hedonics, semantics, pragmatics, as well as affective dimensions (Cariani & Micheyl, 2012).

Appendix B
The latency and amplitude of AEPs across changes in sound level

The figures in this appendix are all due to Picton (2010).

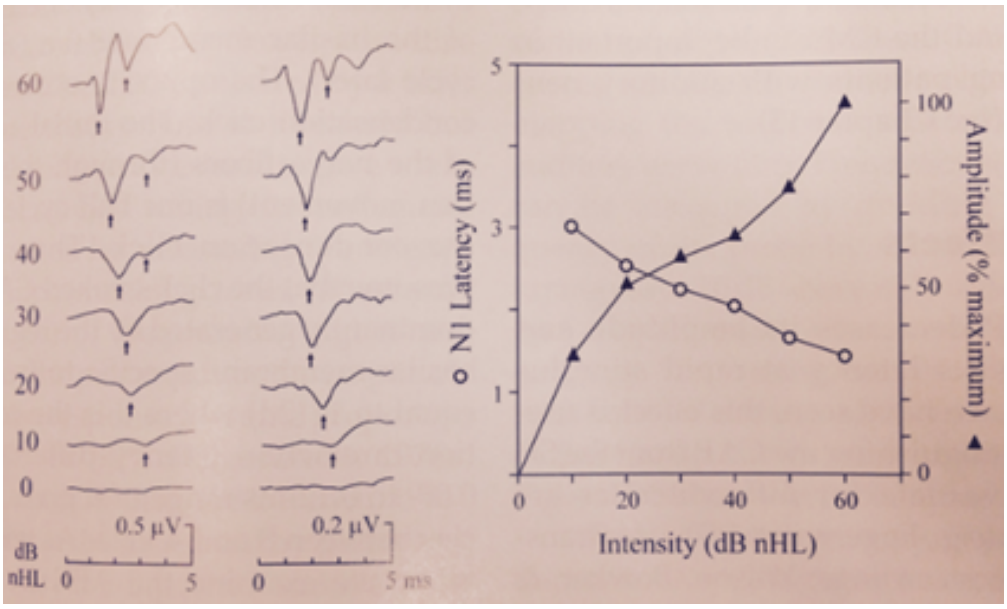

Figure B.1: Compound action potential (CAP) sound level effects. The auditory nerve's CAP response to a 10 s click stimulus presented in alternating polarity at a rate of 5/s at varying sound levels. On the left, the responses of two subjects are shown in 10 dB steps from 0-60 dB nHL. Two different peaks (N1 and N2) become visible at higher levels (marked here by arrows), but only N1 peaks are visible at the lower levels. On the right, the average amplitudes (triangles) and latencies (circles) are plotted as a function of sound level, and indicate a trend of increasing amplitude and decreasing latency with increasing sound level. Reprinted with permission from Human Auditory Evoked Potentials (pg. 203), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

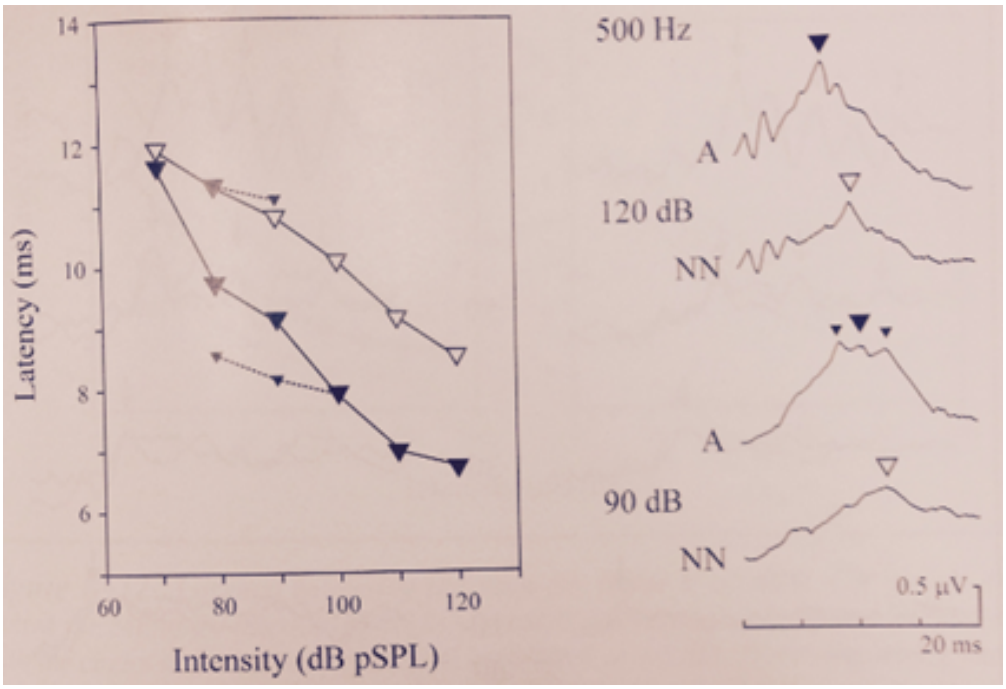

Figure B.2: Auditory brainstem response evoked by 500 Hz tones at various levels. The average brainstem responses of eight normal subjects to a brief 500 Hz tone presented alone (open triangle) and in notched noise (filled triangle) at a rate of 40/s are shown on the left. On the right, the responses of a single subject are shown. Notice that the responses are earlier and larger at the higher 120 dB level in comparison to the 90 dB level in both recording conditions. The addition of notched noise tends to increase latency and decrease amplitude relative to the condition wherein the tone is presented alone, and there are fewer distinguishable peaks in the notched noise condition. Reprinted with permission from Human Auditory Evoked Potentials (pg. 229), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

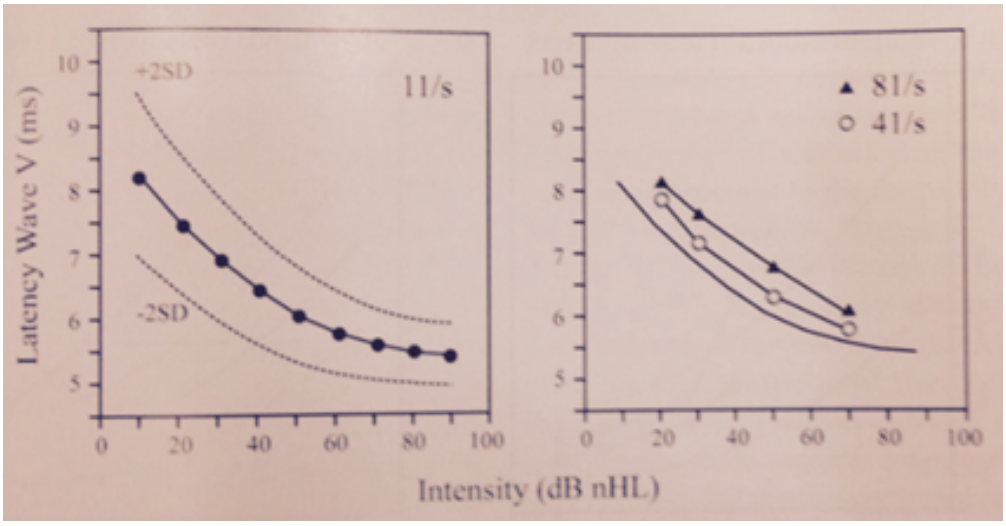

Figure B.3: Latency intensity function for wave V of the auditory brainstem response. The plot on the left shows the change in mean latency with increasing sound level for a click presented at a rate of 11/s. The plot on the right shows the effects of increasing the presentation rate to 41/s or 81/s. In all cases, there is a noticeable trend of decreasing latency with increasing sound level. Reprinted with permission from Human Auditory Evoked Potentials (pg. 231), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

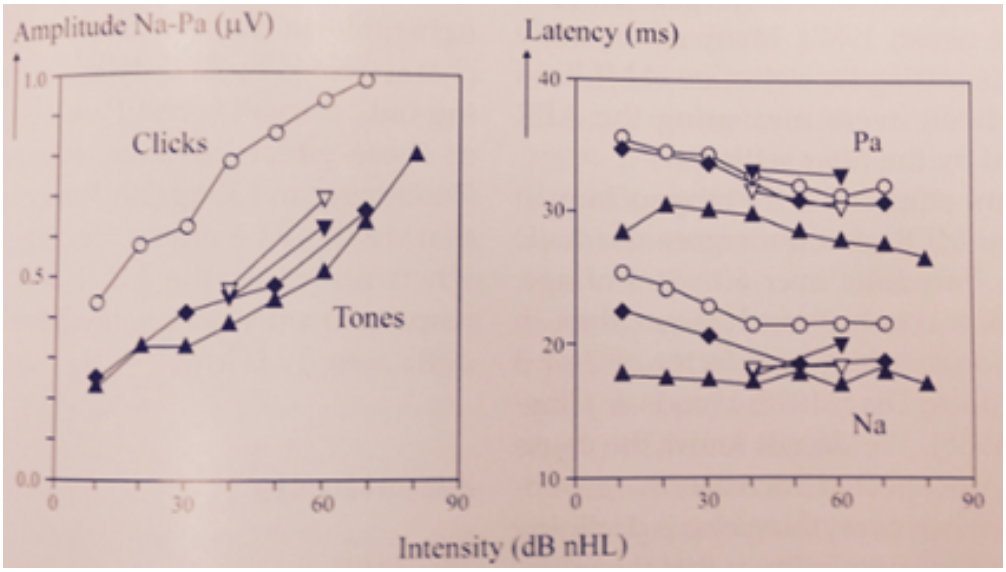

Figure B.4: Stimulus response relations of the middle latency response. On the left, the amplitude of the Na-Pa complex as observed in multiple studies is plotted with respect to stimulus intensity (dB nHL). On the right, the latency of the Na and Pa components are plotted as a function of stimulus intensity. In both cases, the familiar progression of increasing amplitude and decreasing latency as a function of increasing sound level is observed. Reprinted with permission from Human Auditory Evoked Potentials (pg. 253), by T. Picton, San Diego, CA. Copyright 2010 by Plural Publishing Inc.

Appendix C
FFR recording recommendations

Table C.1: Recommended stimulus and presentation parameters. Reprinted with permission from "Auditory Brain Stem Response to Complex Sounds: A Tutorial", by E. Skoe and N. Kraus, 2010, Ear and Hearing, 31, p. 305, 313. Copyright 2010 by Lippincott Williams & Wilkins.

Parameter	Recommendation	Rationale/Comments
Stimulus		
Type	Speech, music, nonspeech vocal sounds, environmental sounds, etc.	Examine how behaviorally relevant sounds are turned into neural code
Characteristics		
Transient	Well-defined temporal features such as strong attacks and amplitude bursts	Maximize transient responses
Sustained	F_0 <300 Hz	Maximize sustained responses
Creation	Natural, synthetic, or hybrid	cABR stimuli can be created with many different software packages
Duration	Short: 40–100 msec	Minimizes recording time
	Long: >100 msec	Maximizes naturalness
Stimulus Presentation		
Intensity	Well above hearing threshold: 60–80 dB SPL	Stimuli should be precisely calibrated before each test session using a sound level meter
Monaural Stimulation	Separate norms should be collected for each ear	Monaural is preferred for children
Binaural Stimulation	Maximizes response characteristics	Binaural is more realistic than monaural
Transducer	Magnetically shielded ear inserts	Minimizes stimulus artifact
Rate and ISI	Rate: dependent on stimulus duration	See Table 3 for recording-based issues that impact rate and ISI decisions
	ISI: ≥30% of stimulus duration	
Presentation Software	Perform thorough testing to ensure precise, nonjittered stimulus presentation	Because of the temporal sensitivity of the cABR, a small amount of jitter will spoil the response
Electrode placement	Vertical montage (active: Cz; reference: earlobe(s); ground: forehead)	For rostral brain stem recordings; a horizontal montage is used for recording from more peripheral structures
Sampling rate	6000–20000 Hz	Better temporal precision with higher sampling rates
Filtering	Low-pass cutoff: 2000–3000 Hz	More defined transient peaks
	High-pass cutoff: 30–100 Hz	Depends on spectral characteristics of stimulus
	If possible, collect cABR with open filters (1–3000 Hz) and band-pass filter off-line using digital filters	Digital filters minimize temporal phase shifts
Signal averaging	2 or more subaverages of 2000–3000 sweeps	Determine response replicability
		Spectral-domain averaging will increase spectral estimates and require fewer sweeps
Averaging window	Begin 10–50 msec before stimulus onset	An adequate sample of the baseline is needed to determine whether a particular response peak is above the noise floor
		For running window analysis, the pre-stimulus time window should be greater than or equal to the duration of the analysis window
	Extend 10–50 msec after stimulus onset	Neural activity should return to baseline
Simultaneous cABR-cortical response recording	Only if large files can be accommodated and longer sessions are appropriate	
Minimizing artifacts	Passive collection protocol	Minimizes myogenic artifacts
	Electromagnetically shielded insert ear phones	Minimize stimulus artifact
	Both stimulus polarities	Enables adding of responses to minimize both stimulus artifact and cochlear microphonic
	Use electrically shielded test booth	Minimizes electrical artifact
	Project movie into test booth	
	Artifact rejection criterion: >20 μV	Exclude trials exceeding typical neural response size; criterion depends on high-pass filter setting

Appendix D
The effects of sound level on the speech-evoked FFR

The following three figures are due to a study by Krishnan (1999) on the spectral frequency following response (sFFR) to two-tone approximations of steady-state vowels presented at four different levels.

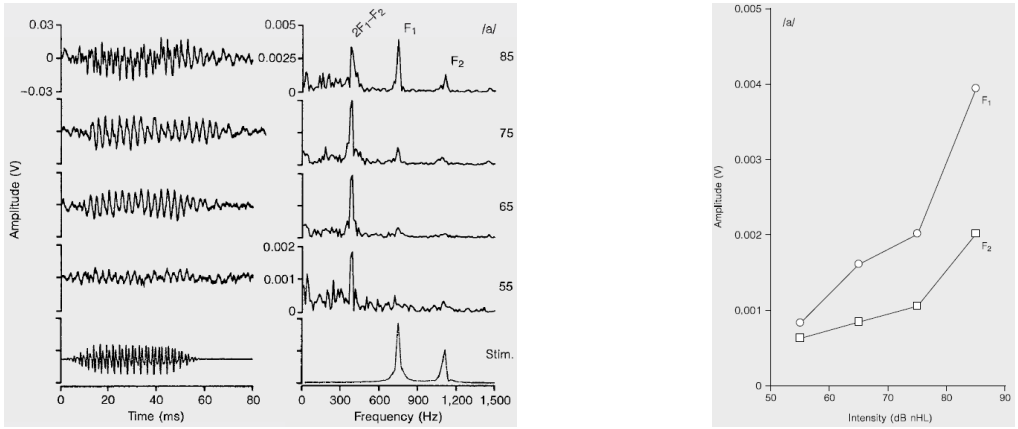

Figure D.1: /a/ vowel mean sFFR waveforms and spectra (left) and the amplitude growth of the first (F1) and second (F2) formants (right). On the left, the waveform and spectra of the stimulus are presented along the bottom row. All levels are in dB nHL. Reprinted with permission from "Human Frequency-Following Responses to Two-Tone Approximations of Steady-State Vowels", by A. Krishnan, 1999, Audiology & Neuro-Otology, 4, p. 99, 100. Copyright 1999 by S. Karger AG.

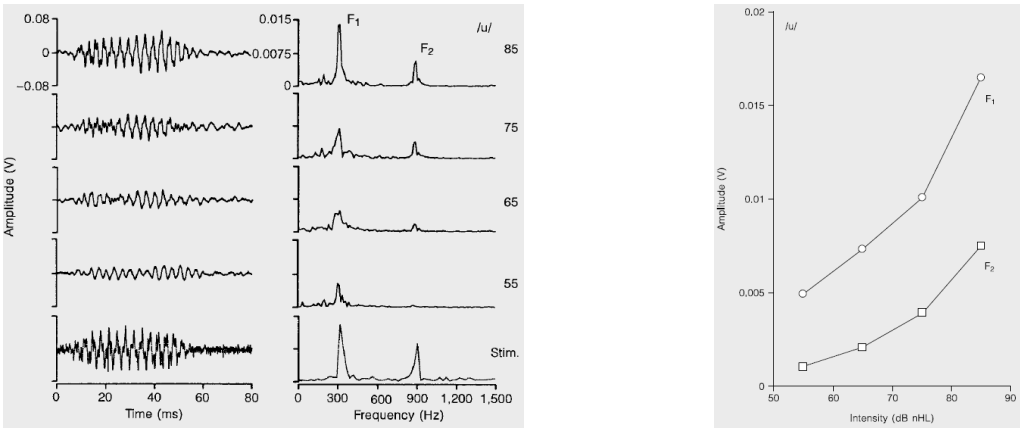

Figure D.2: /u/ vowel mean sFFR and waveforms and spectra (left) and the amplitude growth of the first (F1) and second (F2) formants (right). On the left, the waveform and spectra of the stimulus are presented along the bottom row. All levels are in dB nHL. Reprinted with permission from "Human Frequency-Following Responses to Two-Tone Approximations of Steady-State Vowels", by A. Krishnan, 1999, Audiology & Neuro-Otology, 4, p. 97,98. Copyright 1999 by S. Karger AG.

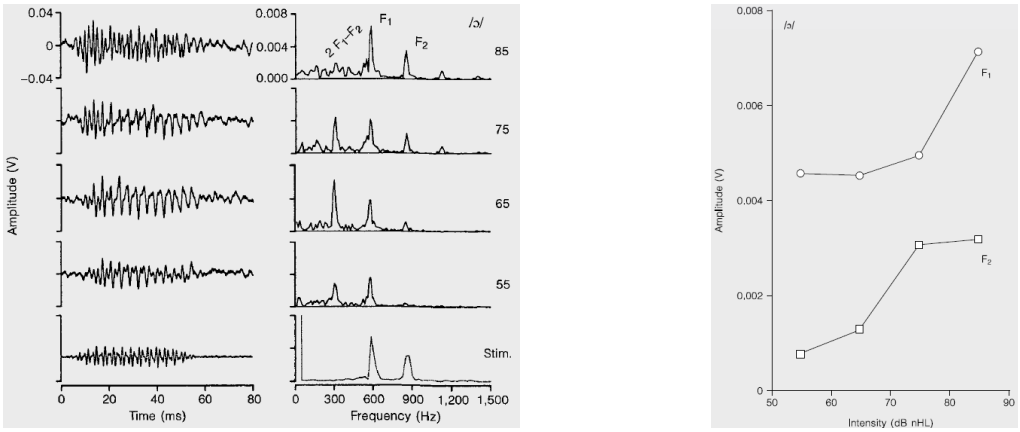

Figure D.3: /ɔ/ vowel mean sFFR and waveforms and spectra (left) and the amplitude growth of the first (F1) and second (F2) formants (right). On the left, the waveform and spectra of the stimulus are presented along the bottom row. All levels are in dB nHL. Reprinted with permission from "Human Frequency-Following Responses to Two-Tone Approximations of Steady-State Vowels", by A. Krishnan, 1999, Audiology & Neuro-Otology, 4, p. 98, 99. Copyright 1999 by S. Karger AG.

The following three figures are due to a study by Krishnan (2002) on the spectral frequency following response (sFFR) to three steady-state synthetic vowels presented at four different levels.

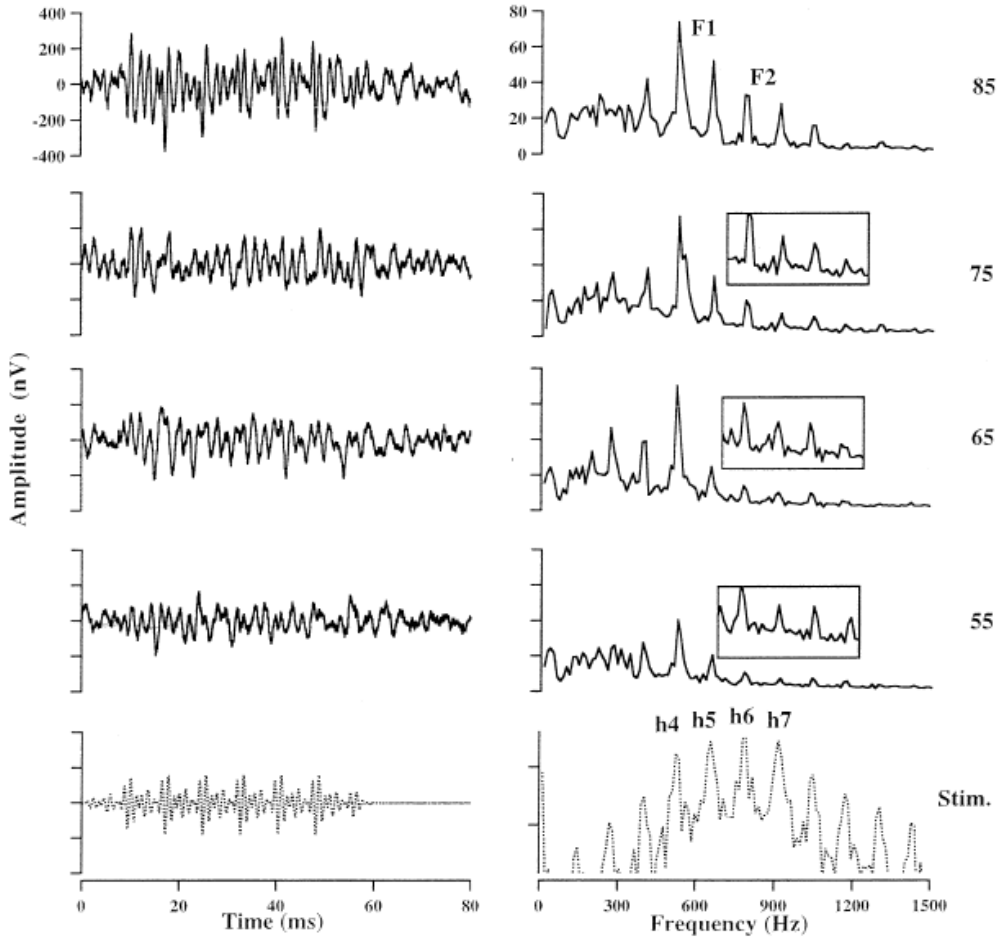

Figure D.4: /ɔ/ vowel grand average sFFR waveforms (left) and spectra (right) plotted for all four level conditions (55, 65, 75, and 85 dB nHL). The first (F1) and second (F2) formants are identified, with the insets zoomed in to show the F2 harmonic peaks more clearly. Reprinted with permission from "Human frequency-following responses: representation of steady-state synthetic vowels", by A. Krishnan, Hearing Research, 166, p. 196. Copyright 2002 Elsevier Science B.V.

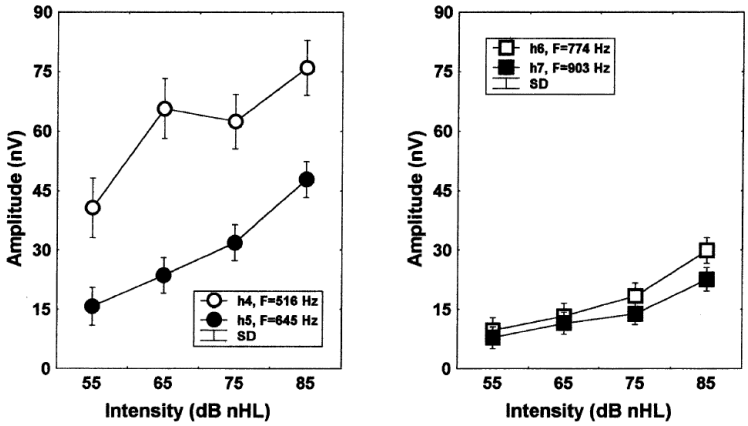

Figure D.5: /ɔ/ vowel mean sFFR amplitudes for the first (left) and second (right) formant harmonics. The error bars indicate ±1 standard deviation. Reprinted with permission from "Human frequency-following responses: representation of steady-state synthetic vowels", by A. Krishnan, Hearing Research, 166, p. 196. Copyright 2002 Elsevier Science B.V.

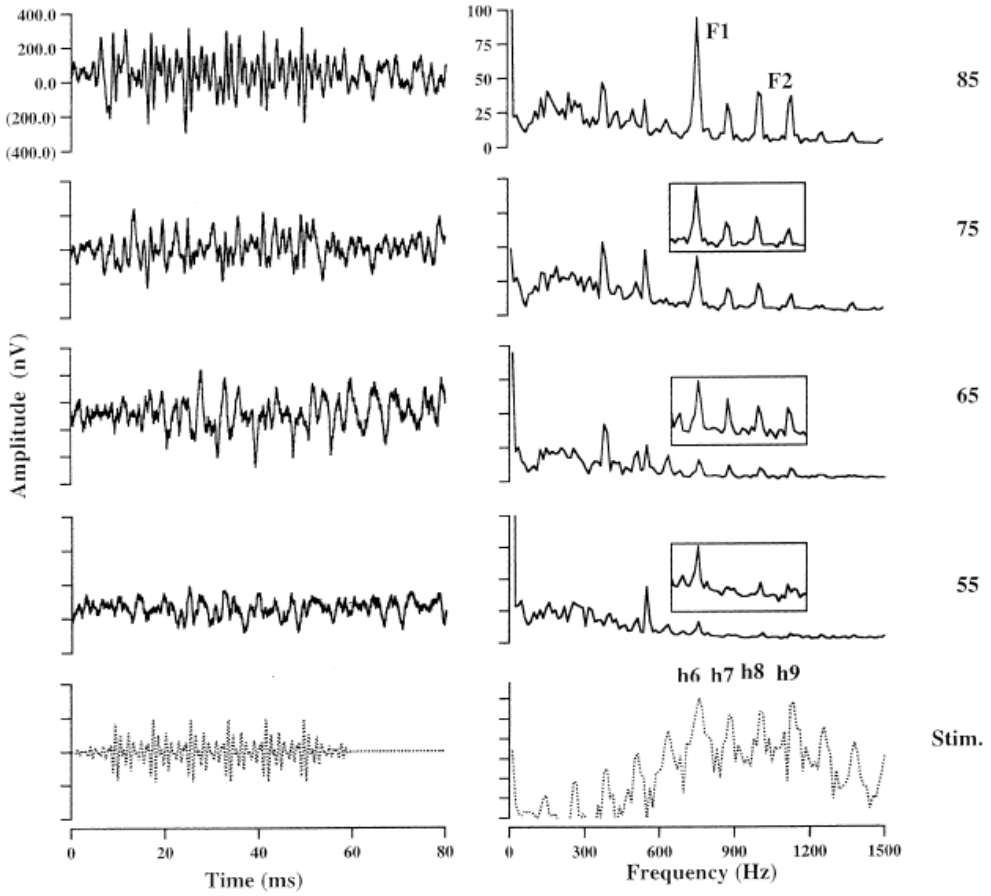

Figure D.6: /a/ vowel grand average sFFR waveforms (left) and spectra (right) plotted for all four level conditions (55, 65, 75, and 85 dB nHL). The first (F1) and second (F2) formants are identified, with the insets zoomed in to show the F2 harmonic peaks more clearly. Reprinted with permission from "Human frequency-following responses: representation of steady-state synthetic vowels", by A. Krishnan, Hearing Research, 166, p. 197. Copyright 2002 Elsevier Science B.V..

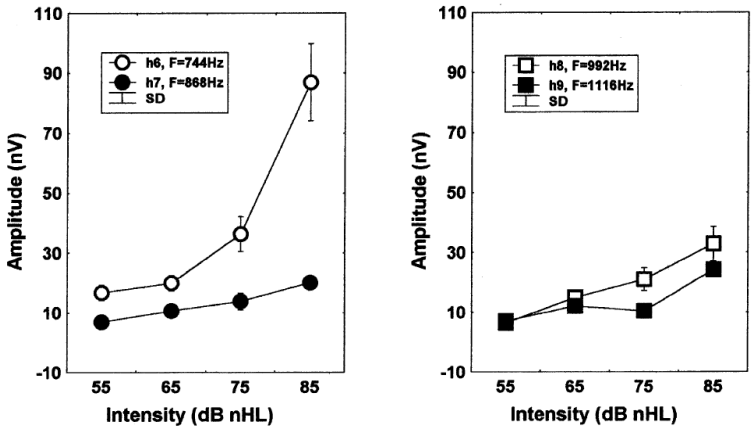

Figure D.7: /a/ vowel mean sFFR amplitudes for the first (left) and second (right) formant harmonics. The error bars indicate ±1 standard deviation. Reprinted with permission from "Human frequency-following responses: representation of steady-state synthetic vowels", by A. Krishnan, Hearing Research, 166, p. 198. Copyright 2002 Elsevier Science B.V.

Appendix E
The Holm-Bonferroni method

The Holm-Bonferroni method is used to control the family-wise error rate when statistical testing involves multiple comparisons (Holm, 1979; Park et al., 2009). This method is preferred to the more widely-known Bonferroni method since it is more powerful (i.e. it decreases the type II error rate), while maintaining the same false positive (type I) error rate, α. Since the overly-conservative Bonferroni method adjusts the significance level by dividing α by the total number of dependent comparisons (e.g. standard t-tests) to be made, it becomes excessively punitive in some sense as the number of comparisons grows in that the risk of failing to properly reject a false null-hypothesis increases.

In order to counteract this loss of power, the Holm-Bonferroni method makes use of the unadjusted p-values, ordered from least to greatest, and compares them to nominal alpha levels of $\frac{\alpha}{m}$ to α, namely $\frac{\alpha}{m}, \frac{\alpha}{m-1}, ..., \frac{\alpha}{2}, \frac{\alpha}{1}$, where m represents the number of pairwise comparisons being performed. The hypotheses are tested sequentially, beginning with the smallest p-value, and the process stops whenever a non-significant result is encountered. At this point, the remaining untested hypotheses - if indeed there are any - are considered to be non-significant. An example with $m = 6$ pairwise comparisons is provided in Table E.1 below.

Here, hypothetical p-values are given, ordered from least to greatest, along with the adjusted values determined by the Holm-Bonferroni method, as well as the conventional Bonferroni method. Clearly, the Holm-Bonferroni's greater power is exhibited here, with four of the six pairwise comparisons attaining statistical significance in contrast to the single significant result obtained via the conventional Bonferroni method. Since the fifth pairwise

Table E.1: The Holm-Bonferroni method versus the Bonferroni method. Both methods maintain the same family-wise error rate. Here $\alpha = .05$.

p-value	**Holm-Bonferroni sig. level**	**sig.**	**Bonferroni sig. level**	**sig.**
.0071	.0083	**Y**	.0083	**Y**
.0092	.0100	**Y**	.0083	N
.0133	.0125	**Y**	.0083	N
.0156	.0167	**Y**	.0083	N
.0307	.0250	N	.0083	N
.0410	.0500	N	.0083	N

comparison failed to achieve significance, the Holm-Bonferroni method was stopped, and the null-hypothesis of the sixth pairwise comparison was not rejected despite the fact that the associated p-value of $p = .0410$ is less than the adjusted criteria of .0500.

Appendix F
The phonemes of American English

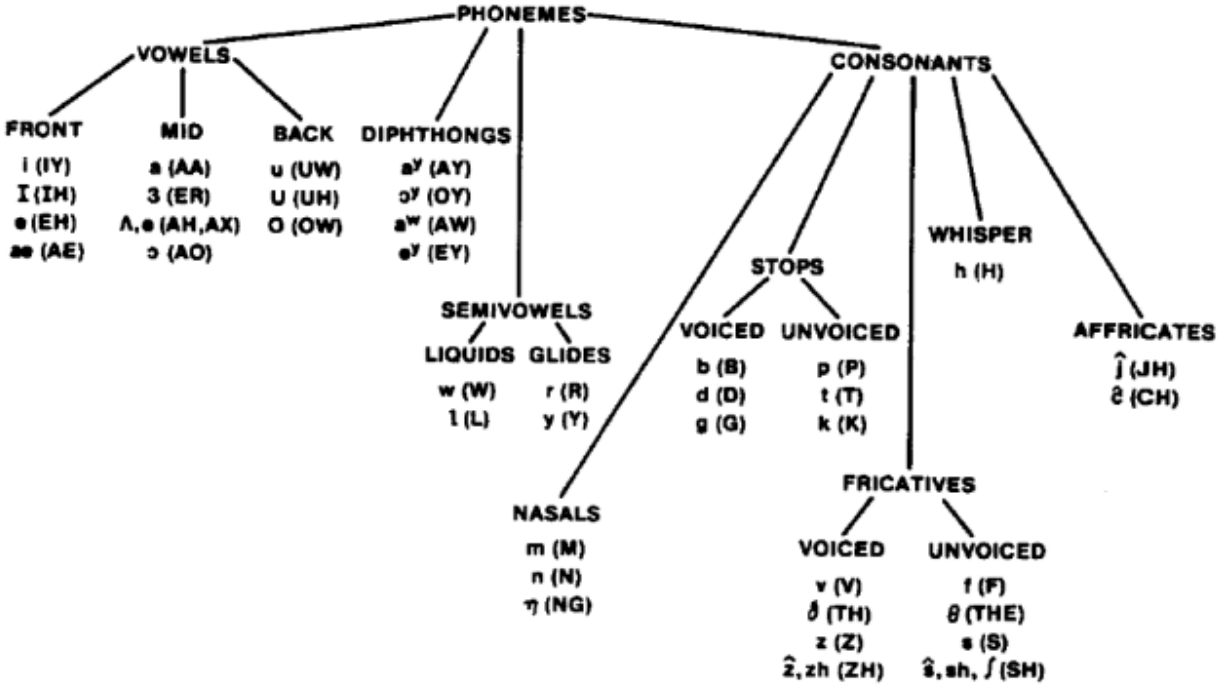

Figure F.1: The phonemes of American English. Reprinted from Fundamentals of Speech Recognition (p. 25), by L. Rabiner and B.-H. Juang, 1993, Englewood Cliffs, NJ: PTR Prentice Hall, Inc. Copyright 1993 by Prentice Hall, Inc.

Appendix G
Test-retest stability of the FFR

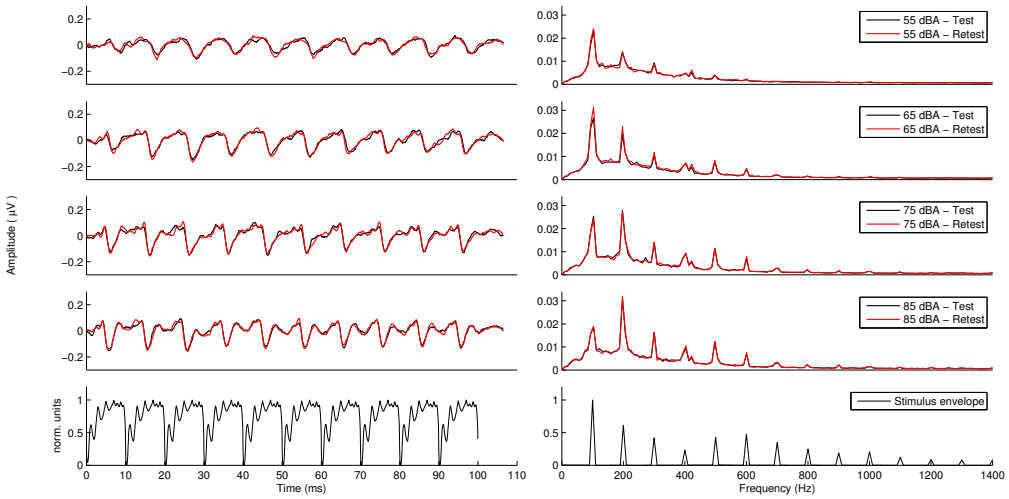

Figure G.1: Test-retest reliability of the 100 ms /a/ vowel eFFR. The test session grand-average (black) is nearly obscured by the retest session grand-average (red).

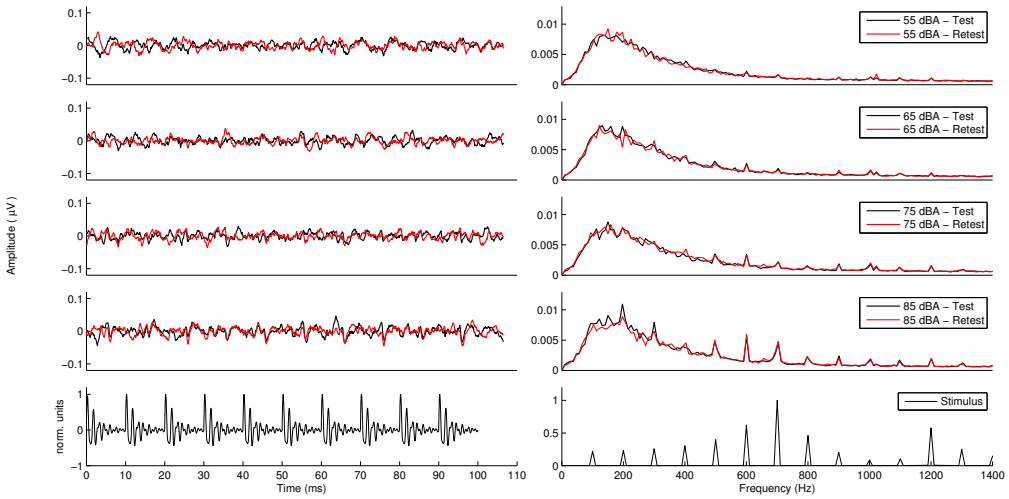

Figure G.2: Test-retest reliability of the 100 ms /a/ vowel sFFR. The test session grand-average (black) shows significant overlap with the retest session grand-average (red) in the area of the first and second formants in the amplitude spectra (right), however the corresponding waveforms (left) appear to be less correlated by comparison.

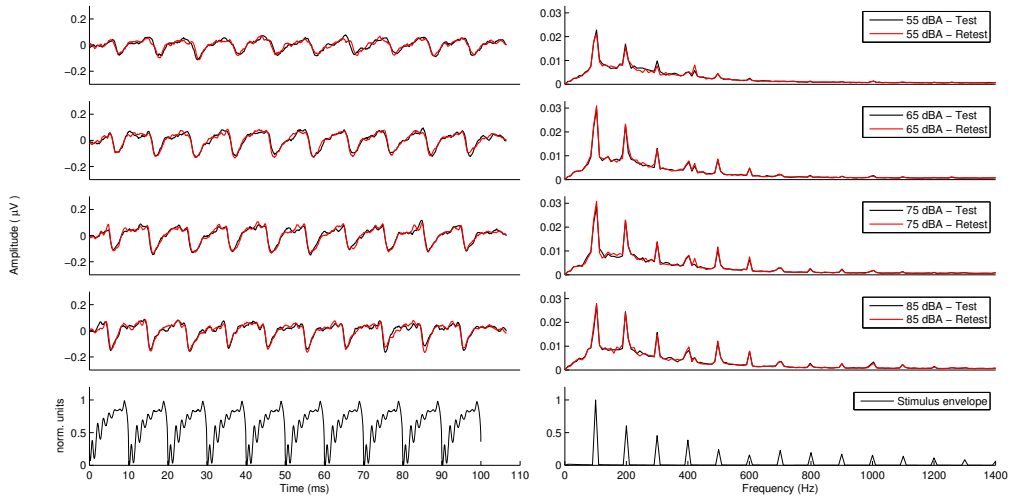

Figure G.3: Test-retest reliability of the 100 ms /U/ vowel eFFR. The test session grand-average (black) is nearly obscured by the retest session grand-average (red).

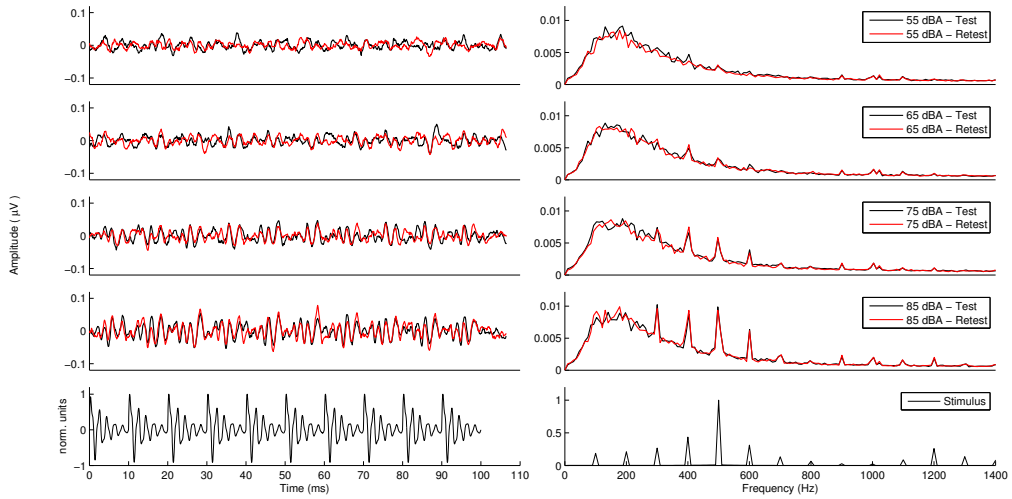

Figure G.4: Test-retest reliability of the 100 ms /U/ vowel sFFR. The test session grand-average (black) shows significant overlap with the retest session grand-average (red) in the area of the first and second formants in the amplitude spectra (right), however the corresponding waveforms (left) appear to be less correlated by comparison.

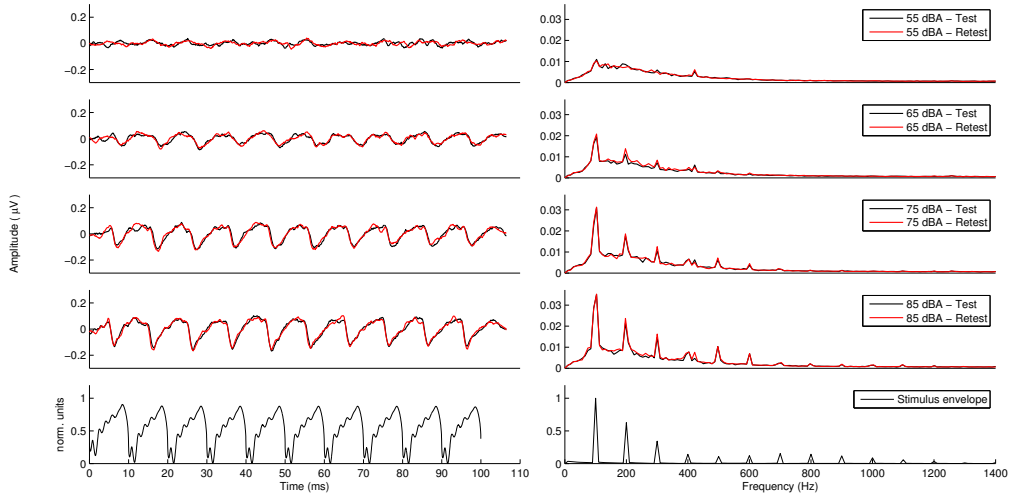

Figure G.5: Test-retest reliability of the 100 ms /u/ vowel eFFR. The test session grand-average (black) is nearly obscured by the retest session grand-average (red).

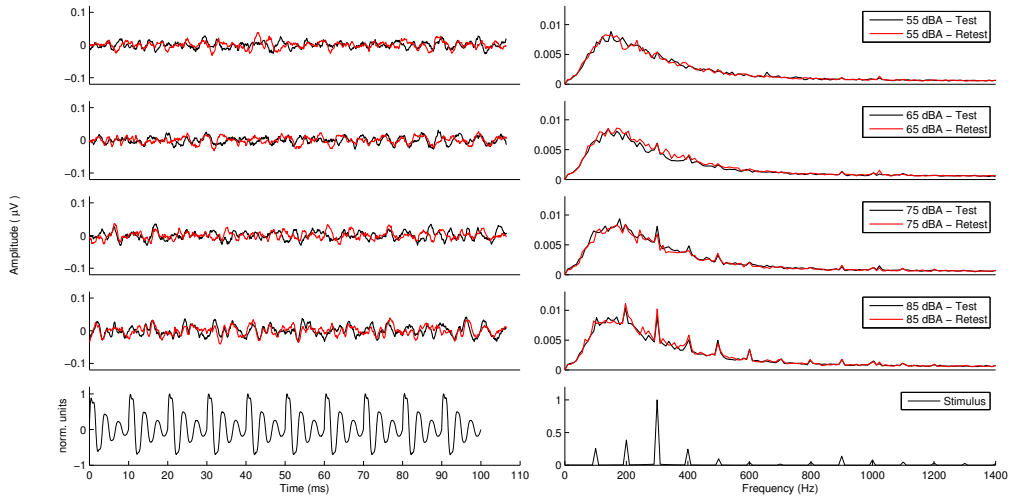

Figure G.6: Test-retest reliability of the 100 ms /u/ vowel sFFR. The test session grand-average (black) shows significant overlap with the retest session grand-average (red) in the area of the first and second formants in the amplitude spectra (right), however the corresponding waveforms (left) appear to be less correlated by comparison.

Appendix H
Additional ML results

The following tables and figures represent additional machine learning results, as referenced in Chapter 5.

Table H.1: The results of the classification performance on the vowel-level task obtained using F0 and its harmonics as features obtained from the combined subaverages. The feature sets used include the use of the eFFR and sFFR harmonics on their own and in combination. The best performance in each instance is indicated in bold.

	Classification accuracy (%) per feature set			
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	6.25	**35.2**	**25.5**	**17.3**
SVM	6.25	18.0	16.4	13.2
LR	6.25	18.4	16.3	12.7
KNN	6.25	28.2 (k=1)	22.8 (k=1)	13.7 (k=37)
DT	6.25	22.9	17.2	13.8
RF	6.25	25.7	21.9	15.3

Table H.2: The results of the classification performance on the vowel task obtained using F0 and its harmonics as features obtained from the combined subaverages. The feature sets used include the use of the eFFR and sFFR harmonics on their own and in combination. The best performance in each instance is indicated in bold.

	Classification accuracy (%) per feature set			
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	25	**58.6**	**47.1**	51.7
SVM	25	39.9	38.0	41.1
LR	25	38.6	36.6	41.2
KNN	25	44.3 (k=1)	42.4 (k=14)	71.5 (k=1)
DT	25	44.8	37.6	70.3
RF	25	54.2	43.1	**73.8**

Table H.3: The results of the classification performance on the level task obtained using F0 and its harmonics as features obtained from the combined subaverages. The feature sets used include the use of the eFFR and sFFR harmonics on their own and in combination. The best performance in each instance is indicated in bold.

	Classification accuracy (%) per feature set			
Algo.	**Chance (null set)**	**eFFR & sFFR**	**eFFR only**	**sFFR only**
LDA	25	**55.1**	47.4	**36.6**
SVM	25	43.5	41.3	34.8
LR	25	43.4	41.6	34.2
KNN	25	52.8 (k=13)	**49.0 (k=15)**	34.1 (k=37)
DT	25	45.9	42.3	34.2
RF	25	54.6	45.2	32.7

Table H.4: Classification accuracy results obtained on the all-subaverages eFFR and sFFR harmonic features with additional amplitude spectra from the two bins on either side of each harmonic (i.e. noise). The results are comparable though slightly worse than those obtained without the additional 'noise' features.

	Classification accuracy (%) per predictive task		
Algo.	**Vowel & level (Chance: 6.25)**	**Vowel only (Chance: 25)**	**Level only (Chance: 25)**
LDA	**28.0**	**50.8**	**51.9**
SVM	19.1	38.7	43.4
LR	17.4	36.9	40.9
KNN	22.5 (k=5)	41.2 (k=38)	50.3 (k=20)
DT	18.3	42.4	43.3
RF	18.7	41.4	49.9

Table H.5: Classification accuracy results obtained on the vowel-level, vowel, and level classification tasks using the leave-one-out (LOO) cross-validation technique and the combination of eFFR and sFFR harmonic features. The best classifier performance in each instance is indicated in bold.

	Classification accuracy (%) per predictive task		
Algo.	**Vowel & level (Chance: 6.25)**	**Vowel only (Chance: 25)**	**Level only (Chance: 25)**
LDA	**33.9**	**55.6**	**53.8**
SVM	16.1	38.3	44.2
LR	6.5	30.5	39.2
KNN	28.0 (k=1)	46.0 (k=1)	53.1 (k=37)
DT	23.4	47.3	44.4
RF	27.1	50.2	53.2

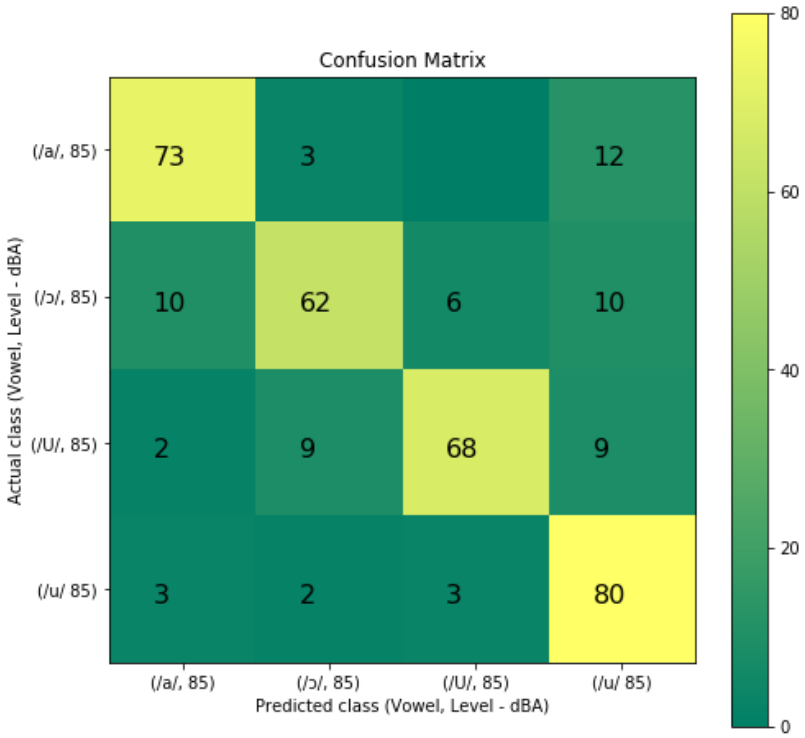

Figure H.1: Vowel classification accuracy of an SVM trained on the eFFR time series data from the 85 dBA data level-condition only. The classification accuracy of 80.8% is comparable to the 83.3% obtained in the study by Sadeghian et. al (2012) that classified vowels at a static level of 80.5 dB SPL.

Appendix I
Ethics approval

The informed consent forms appended below, one in English and one in French, were utilized for Studies 1 and 2 in accordance with the University of Ottawa's ethics approval for the research project led by Dr. Dajani entitled: "Measurement and Applications of Speech-evoked Auditory Brainstem Responses (sABRs)" (file no. H05-14-22B). Both Dr. Giguère and myself, Brian Heffernan, were listed as co-investigators on this project throughout the duration of the aforementioned studies.

Informed Consent Form

Measurement of Speech-Evoked Potentials in Normal Hearing Subjects

You are invited to participate in this study. This informed consent form is to make sure that you understand the purpose of this study and the nature of your involvement. The research project is being conducted by Dr. Hilmi Dajani, Associate Professor at the School of Electrical Engineering and Computer Science ([redacted]). Data from this study may be used by Mr. Rida Al-Osman ([redacted]) or Mr. Brian Heffernan ([redacted]), for their thesis projects, or by their thesis co-supervisor Dr. Christian Guiguère, Full Professor at the School of Rehabilitation Sciences ([redacted]).

Purpose of the study: The purpose of this study is to investigate the brain's electrical response to speech stimuli. This electrical signal, called the auditory-evoked potential, allows clinicians to diagnose hearing problems and localize their origin in the brain. The sounds that clinicians most commonly use for this test are brief clicks or other artificial sounds. These types of sounds do not provide sufficient information to assess the ability to hear speech. As a result, we are interested in developing tests that record the brain's electrical response to speech, or the Speech-Evoked Potential, with a diverse set of speech samples and in different acoustic environments. The knowledge gained is expected to help in the development of better diagnostic tests of hearing loss and tools for evaluating hearing aids.

Procedure: The study is performed in the Audiology Laboratory, 1st floor, Roger-Guindon Hall, 451 Smyth Rd., Ottawa, ON K1H 8M5. Data is collected on a single day and the entire data collection protocol lasts about two hours.

The following is the step by step procedure that will be followed:

1. The participant's ears will first be examined using an otoscope to check for outer and middle ear problems.
2. The hearing of the participant will be tested with pure tones at 500, 1000, 2000 and 4000 Hz.
3. Participants will be asked to sit in a comfortable chair inside a sound-proof room, and will be asked to minimize movement during the recordings.
4. The vertex (at centre of the top of the scalp) and earlobes will be cleaned with a mild abrasive. Then, an electrode filled with conductive gel will be placed on the vertex and two other electrodes will be placed on the earlobes to measure the brain's response to sounds. Foam earphones will be inserted into the participant's ear canals. The earphones will deliver the sound at a level calibrated at less than or equal to 85 dB SPL. This level is well within the

acceptable safe range for sounds, and well below Canadian occupational exposure limits.

Subject description: Adults with normal hearing and without any external or middle ear disease will be included.

Risks to participating: There is no danger or risk to health associated with this study. All procedures have been pre-tested and they have been used routinely for many years in hospitals, clinics, and laboratories.

Withdrawing from the study: Your participation in this study is voluntary. You may withdraw from the study at any time, by verbally informing any of the investigators, even after signing this form, or even after data collection has ended. There will be no consequences following this action, and the collected data will be deleted.

Compensation: You will receive $10 for each hour or part of an hour as monetary compensation.

Confidentiality: Any information about you collected during this study will be kept strictly confidential. Your name will not be associated with the collected data in any way. While the results will appear in students' dissertations and may also be published, you will not be identified.

The data collection is conducted by Dr. Dajani or his graduate student. The records will be kept on computer files. The files will be password-protected. The data will be kept in Dr. Dajani's office when not in use and will be conserved for 15 years, after which the computer files will be deleted.

Benefits of the Study

Hearing impairment is an important public health problem. This study will help develop better technologies to measure hearing loss.

In closing: With your participation, you will be given a copy of this consent form. At the conclusion of the study, should you wish, you will be provided with a summary of the results. You may ask questions at any time, even after signing this consent form.

If you have any concerns with regards to the ethical conduct of the study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5, tel.: 613-562-5841, email: ethics@uottawa.ca.

Signatures: I have read the above description of the study and understand the conditions of participation. My signature indicates that I agree to participate in the study.

Name of the participant (please print):

_____

Participant’s signature:

_____ Date: _____

Researcher's signature:

_____ Date: _____

Formulaire de consentement informé

Mesure des potentiels évoqués par la parole pour les sujets avec une audition normale

Vous êtes invités à participer à cette étude. Ce formulaire de consentement informé permet de s'assurer que vous comprenez le but de cette étude et la nature de votre engagement. Le projet de recherche est dirigé par Dr. Hilmi Dajani, professeur agrégé à l'école d'ingénierie et de technologie de l'information de l'Université d'Ottawa ([redacted]). Les données de cette étude pourront être utilisées par M. Rida Al-Osman ([redacted]) ou M. Brian Heffernan ([redacted]) pour leurs projets de thèse, ou par leur codirecteur de thèse Dr Christian Guiguère, professeur titulaire à l'École des sciences de la réadaptation ([redacted]).

But de l'étude: Le but de cette étude est d'étudier la réponse électrique du cerveau au stimulus de la parole. Ce signal électrique, appelé le potentiel évoqué auditif, permet à des cliniciens de diagnostiquer des problèmes d'audition et de localiser leur origine dans le cerveau. Les sons le plus couramment utilisés par les cliniciens sont des clics assez courts ou bien d'autres sons artificiels. Ces types de sons ne fournissent pas d'informations suffisantes pour évaluer la capacité de percevoir la parole. En conséquence, nous souhaitons développer des tests qui enregistrent la réponse électrique du cerveau à la parole, autrement le potentiel évoqué par la parole, avec un ensemble divers d'échantillons de la parole et dans différents environnements acoustiques. On s'attend à ce que les connaissances acquises aident dans le développement de meilleurs tests diagnostic de perte d'audition et d'outils pour évaluer des prothèses auditives.

Procédure: L'étude est menée dans le laboratoire d'audiologie, 1er étage, pavillon Roger-Guindon, 451 chemin Smyth, Ottawa, ON K1H 8M5. Les données sont collectées sur une seule journée, et le protocole entier de collecte de données dure environ deux heures.

Ce qui suit est la procédure qui sera suivie:
1. Les oreilles du participant seront d'abord examinées à l'aide d'un otoscope pour vérifier l'absence des problèmes de l'oreille externe et moyenne.
2. L'audition du participant sera testée avec des tons purs à 500, 1000, 2000 et 4000 Hz.
3. Les participants seront invités à s'asseoir dans un fauteuil confortable dans une pièce insonorisée, et seront invités à minimiser le mouvement pendant les enregistrements.
4. Le sommet (au centre de la partie supérieure du cuir chevelu) et les lobes de l'oreille seront nettoyés avec un abrasif doux. Puis, une électrode remplie de gel conducteur sera placée sur le sommet et deux autres électrodes seront

mises sur les lobes de l'oreille pour mesurer la réponse du cerveau aux sons. Des écouteurs en mousse seront insérés dans les canaux auditifs du participant. Les écouteurs fourniront le son à un niveau calibré inférieur ou égal à 85 dB SPL. Ce niveau d'intensité est tout à fait en conformité avec les limites sécuritaires acceptables pour les sons, et bien en dessous des limites d'exposition professionnelles canadiennes.

Description des sujets: Des adultes avec une audition normale et sans problème de l'oreille externe et moyenne seront inclus dans l'expérience.

Les risques encourus par la participation: Il n'y a pas de danger ou de risque pour la santé associée à cette étude. Toutes les procédures ont été prétestées et elles sont utilisées régulièrement, et ce, depuis plusieurs années par de nombreux hôpitaux, cliniques et laboratoires.

Le retrait de la participation à l'étude: Votre participation à cette étude est volontaire. Vous pouvez en tout temps vous retirer de l'étude, en informant verbalement tous chercheurs ou cochercheurs, et ce, même après avoir signé ce formulaire, ou même après la collecte des données est terminée. Il n'y aura aucune conséquence suivant cette action, et les données collectées seront effacées.

Compensation monétaire: Vous recevrez 10 $ pour chaque heure ou fraction d'heure à titre de compensation monétaire.

Confidentialité: Toutes informations recueillies à votre sujet durant l'étude seront conservées de manière strictement confidentielle. Votre nom ne sera en aucun cas associé avec la collecte de données. Alors que les résultats pourront soit figurer dans les dissertations des étudiants impliqués ou soit dans les articles scientifiques, vous ne serez pas identifié.

La collecte de données est effectuée par Dr. Dajani ou bien son étudiant diplômé. Ces données seront conservées dans des fichiers d'ordinateur et seront protégées par un mot de passe. Les données seront gardées dans le bureau du Dr. Dajani quand elles ne sont pas utilisées et seront conservées pour une période de 15 ans. Après 15 ans, les fichiers d'ordinateur seront supprimés.

Les bénéfices de l'étude: La déficience auditive est un problème important de santé publique. Cette étude permettra de développer de meilleures technologies pour mesurer la perte d'audition.

Pour conclure: Avec votre participation, vous serez donné une copie de ce formulaire de consentement. À la fin de cette étude, vous pourrez, si vous le désirez, recevoir un résumé des résultats. Vous pouvez poser des questions en tout temps lors de l'étude, même après avoir signé ce formulaire de consentement.

Si vous avez des questions concernant l'éthique de cette étude, vous pouvez contacter Le bureau d'éthique pour les protocoles en recherche à l'université d'Ottawa, pavillon Tabaret, 550 Cumberland, pièce 159, Ottawa, ON, K1N 6N5, tel : 613-562-5841, courriel : ethics@uottawa.ca.

Signatures: J'ai lu la description ci-haut de l'étude et j'ai compris les conditions de participation. Ma signature indique que je consens à participer à l'étude.

Nom du participant (caractère imprimé S.V.P.):_____

Signature du participant : _____ Date : _____

Signature du chercheur: _____ Date: _____

References

Ahn, J. H., Lee, H.-S., Kim, Y.-J., Yoon, T. H., & Chung, J. W. (2007). Comparing pure-tone audiometry and auditory steady state response for the measurement of hearing loss. *Otolaryngology—Head and Neck Surgery*, *136*(6), 966–971.

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1-2), 35–47.

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., … Thai-Van, H. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/in normal-hearing adults. *Clinical Neurophysiology*, *119*(4), 922–933.

Alain, C., Arsenault, J. S., Garami, L., Bidelman, G. M., & Snyder, J. S. (2017). Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Scientific Reports*, *7*, 40790.

Alnahwi, M., AlQudehy, Z. A., et al. (2015). Comparison between frequency transposition and frequency compression hearing aids. *The Egyptian Journal of Otolaryngology*, *31*(1), 10.

Al Osman, R. (2016). *A study of auditory speech processing using brainstem evoked responses under the effects of stressors* (Doctoral dissertation). Université d'Ottawa/University of Ottawa.

Al Osman, R., Giguère, C., & Dajani, H. R. (2016). Effects of stimulus rate and noise on speech-evoked auditory brainstem responses. *Canadian Journal of Speech-Language Pathology & Audiology*, *40*(2).

Al Osman, R., Giguère, C., & Dajani, H. R. (2018). Effects of early-and late-arriving room reflections on the speech-evoked auditory brainstem response. *Journal of the American Academy of Audiology*, *29*(2).

Amieva, H., Ouvrard, C., Giulioli, C., Meillon, C., Rullier, L., & Dartigues, J.-F. (2015). Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: A 25-year study. *Journal of the American Geriatrics Society*, *63*(10), 2099–2104.

Ananthakrishnan, S., Krishnan, A., & Bartlett, E. (2016). Human frequency following response: neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss. *Ear and Hearing*, *37*(2), e91.

Anderson, S. (2017). Clinical translation: Aging, hearing loss, and amplification. In *The frequency-following response* (pp. 267–294). Springer.

Anderson, S., & Kraus, N. (2013). The potential role of the cabr in assessment and management of hearing impairment. *International Journal of Otolaryngology*, *2013*.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Drehobl, S., & Kraus, N. (2013). Effects of hearing loss on the subcortical representation of speech cues. *The Journal of the Acoustical Society of America*, *133*(5), 3030–3038.

Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, *32*(6), 750.

Bainbridge, K. E., & Ramachandran, V. (2014). Hearing aid use among older united states

adults: the national health and nutrition examination survey, 2005–2006 and 2009–2010. *Ear and Hearing*, *35*(3), 289.

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and Language*, *126*(1), 29–48.

Batta, M., Dhir, S. K., Kumar, A., & Singh, K. (2017). Effect of different phases of menstrual cycle on brainstem auditory evoked response. *International Journal of Applied and Basic Medical*, *7*(1), 44.

Beck, D. L., & Nilsson, M. (2013). Speech-in-noise testing: A pragmatic addendum to hearing aid fittings. *Hearing Review*, *16*, 24–26.

Behler, O., & Uppenkamp, S. (2016a). Auditory fmri of sound intensity and loudness for unilateral stimulation. In *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 165–174). Springer.

Behler, O., & Uppenkamp, S. (2016b). The representation of level and loudness in the central auditory system for unilateral stimulation. *NeuroImage*, *139*, 176–188.

Beigi, H. (2011). *Fundamentals of speaker recognition*. Springer Science & Business Media.

Bellier, L., Veuillet, E., Vesson, J.-F., Bouchet, P., Caclin, A., & Thai-Van, H. (2015). Speech auditory brainstem response through hearing aid stimulation. *Hearing Research*, *325*, 49–54.

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., & Shinn-Cunningham, B. G. (2015). Individual differences reveal correlates of hidden hearing deficits. *Journal of Neuroscience*, *35*(5), 2161–2172.

Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2014). Rapid acquisition of auditory subcortical steady state responses using multichannel recordings. *Clinical Neurophysiology*, *125*(9), 1878–1888.

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., & Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Frontiers in Systems Neuroscience*, *8*, 26.

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient abr. *Hearing Research*, *323*, 68–80.

Bidelman, G. M. (2016). Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *The Journal of the Acoustical Society of America*, *140*(4), EL358–EL363.

Bidelman, G. M. (2017). Communicating in challenging environments: Noise and reverberation. In *The frequency-following response* (pp. 193–224). Springer.

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*, *175*, 56–69.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter-and intra-hemispheric cortical processing underlying degraded speech perception. *Neuroimage*, *124*, 581–590.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, *79*, 201–212.

Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014). Age-related changes in the subcortical–cortical encoding and categorical perception of speech. *Neurobiology of Aging*, *35*(11), 2526–2540.

Billings, C. J. (2013). Uses and limitations of electrophysiology with hearing aids. In *Seminars in hearing* (Vol. 34, p. 257).

Billings, C. J., McMillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *Journal of the Association for Research in Otolaryngology*, *14*(6), 891–903.

Billings, C. J., Penman, T. M., McMillan, G. P., & Ellis, E. (2015). Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment & age. *Ear and Hearing*, *36*(6), 710.

Billings, C. J., Tremblay, K. L., & Miller, C. W. (2011). Aided cortical auditory evoked potentials in response to changes in hearing aid gain. *International Journal of Audiology*, *50*(7), 459–467.

Billings, C. J., Tremblay, K. L., Souza, P. E., & Binns, M. A. (2007). Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials. *Audiology and Neurotology*, *12*(4), 234–246.

Billings, G., Piasini, E., Lőrincz, A., Nusser, Z., & Silver, R. A. (2014). Network structure within the cerebellar input layer enables lossless sparse encoding. *Neuron*, *83*(4), 960–974.

Billings Curtis, J., Papesh Melissa, A., Penman Tina, M., Baltzell Lucas, S., & Gallun Frederick, J. (2012). Clinical use of aided cortical auditory evoked potentials as a measure of physiological detection or physiological discrimination. *International Journal of Otolaryngology*, *2012*.

Brummer, N. B., Rawson, J. A., Grimm, M. W., Tupper, P., Miller, A. J., Billings, C. J., & Jacobs, P. G. (2013). Electrode selection and other design considerations for a modular, portable single-channel eeg-augmented hearing aid. In *Neural engineering (ner), 2013 6th international ieee/embs conference on* (pp. 431–434).

Burkard, R. F., Eggermont, J. J., & Don, M. (2007). *Auditory evoked potentials: basic principles and clinical application*. Lippincott Williams & Wilkins.

Camacho, L. F., & Kuvadia, S. (2017, November 17). *Datalogging - hearing aid behavior in the real world.* Retrieved 2018-09-17, from `https://www.audiologyonline.com/articles/datalogging-hearing-aid-behavior-in-20258`

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity*, *6*(4), 147–172.

Cariani, P., & Micheyl, C. (2012). Toward a theory of information processing in auditory cortex. In *The human auditory cortex* (pp. 351–390). Springer.

Carlile, S. (2016, February). *Brain Computer Interfaces and Hearables.* Retrieved 2017-01-10, from `http://www.starkey.com/blog/2016/02/brain-computer-interface-hearables`

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Eneuro*, *2*(4), ENEURO–0004.

Carter, L., Dillon, H., Seymour, J., Seeto, M., & Van Dun, B. (2013). Cortical auditory-evoked potentials (caeps) in adults in response to filtered speech stimuli. *Journal of the American Academy of Audiology*, *24*(9), 807–822.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech:

Neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246.

Chang, H.-W., Dillon, H., Carter, L., Van Dun, B., & Young, S.-T. (2012). The relationship between cortical auditory evoked potential (caep) detection and estimated audibility in infants with sensorineural hearing loss. *International Journal of Audiology*, *51*(9), 663–670.

Chasin, M. (2011). Setting hearing aids differently for different languages. In *Seminars in hearing* (Vol. 32, pp. 182–188).

Chen, Y.-H., de Beeck, M. O., Vanderheyden, L., Carrette, E., Mihajlović, V., Vanstreels, K., … Van Hoof, C. (2014). Soft, comfortable polymer dry electrodes for high quality ecg and eeg recording. *Sensors*, *14*(12), 23758–23780.

Choi, C. T., & Lee, Y.-H. (2012). A review of stimulating strategies for cochlear implants. In *Cochlear implant research updates.* InTech.

Chun, I., Billings, C. J., Miller, C. W., & Tremblay, K. L. (2016). Aided electrophysiology using direct audio input: Effects of amplification and absolute signal level. *American Journal of Audiology*, *25*(1), 14–24.

Coffey, E. B., Colagrosso, E. M., Lehmann, A., Schönwiesner, M., & Zatorre, R. J. (2016). Individual differences in the frequency-following response: relation to pitch perception. *PloS One*, *11*(3), e0152374.

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by meg. *Nature Communications*, *7*, 11070.

Coffey, E. B., Musacchia, G., & Zatorre, R. J. (2017). Cortical correlates of the auditory frequency-following and onset responses: Eeg and fmri evidence. *Journal of Neuroscience*, *37*(4), 830–838.

Contrera, K. J., Betz, J., Deal, J., Choi, J. S., Ayonayon, H. N., Harris, T., … others (2017). Association of hearing impairment and anxiety in older adults. *Journal of Aging and Health*, *29*(1), 172–184.

Dajani, H. R., Heffernan, B. P., & Giguère, C. (2013). Improving hearing aid fitting using the speech-evoked auditory brainstem response. In *Engineering in medicine and biology society (embc), 2013 35th annual international conference of the ieee* (pp. 2812–2815).

Da Silva, I. G. A. (2009). *Objective estimation of loudness growth using tone burst evoked auditory responses* (Doctoral dissertation). Northeastern University.

Dehan, C. P., & Jerger, J. (1990). Analysis of gender differences in the auditory brainstem response. *The Laryngoscope*, *100*(1), 18–24.

de Lafuente, V., & Romo, R. (2005). Neuronal correlates of subjective sensory experience. *Nature Neuroscience*, *8*(12), 1698.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *68*(3), 843–857.

Deller, J., Hansen, J., & Proakis, J. (2000). *Discrete-Time Processing of Speech Signals.* Wiley.

Deng, L., & Platt, J. C. (2014). Ensemble deep learning for speech recognition. In *Fifteenth annual conference of the international speech communication association.*

Deo, N., & Grosh, K. (2004). Two-state model for outer hair cell stiffness and motility. *Biophysical Journal*, *86*(6), 3519–3528.

Dillon, H. (2012). *Hearing aids*. Thieme.

Dolphin, W., & Mountain, D. (1992). The envelope following response: scalp potentials elicited in the mongolian gerbil using sinusoidally am acoustic signals. *Hearing Research*, *58*(1), 70–78.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.

Du, Y., Kong, L., Wang, Q., Wu, X., & Li, L. (2011). Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem". *Neuroscience & Biobehavioral Reviews*, *35*(10), 2046–2057.

Easwar, V., Purcell, D. W., Aiken, S. J., Parsa, V., & Scollie, S. D. (2015a). Effect of stimulus level and bandwidth on speech-evoked envelope following responses in adults with normal hearing. *Ear and Hearing*, *36*(6), 619–634.

Easwar, V., Purcell, D. W., Aiken, S. J., Parsa, V., & Scollie, S. D. (2015b). Evaluation of speech-evoked envelope following responses as an objective aided outcome measure: Effect of stimulus level, bandwidth, and amplification in adults with hearing loss. *Ear and Hearing*, *36*(6), 635–652.

Easwar, V., Purcell, D. W., & Scollie, S. D. (2012). Electroacoustic comparison of hearing aid output of phonemes in running speech versus isolation: Implications for aided cortical auditory evoked potentials testing. *International Journal of Otolaryngology*, *2012*.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, *11*(1), 31–45.

Emara, A. A. Y., & Kolkaila, E. A. (2010). Prediction of loudness growth in subjects with sensorineural hearing loss using auditory steady state response. *The Journal of International Advanced Otology*, *6*(3), 371.

Emmett, S. D., & Francis, H. W. (2015). The socioeconomic impact of hearing loss in us adults. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *36*(3), 545.

Energy harvesting from jaw movement to power hearing aids. (2014, July). Retrieved 2018-09-17, from `https://www.offgridenergyindependence.com/articles/6690/energy-harvesting-from-jaw-movement-to-power-hearing-aids`

Feder, K., Michaud, D., Ramage-Morin, P., McNamee, J., & Beauregard, Y. (2015). Prevalence of hearing loss among canadians aged 20 to 79: Audiometric results from the 2012/2013 canadian health measures survey. *Health Reports*, *26*(7), 18.

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-eeg detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, *14*(3), 036020.

Fiedler, P., Strohmeier, D., Hunold, A., Griebel, S., Mühle, R., Schreiber, M., … others (2016). Modular multipin electrodes for comfortable dry eeg. In *Engineering in medicine and biology society (embc), 2016 ieee 38th annual international conference of the* (pp. 5705–5708).

Flowers, P. F., Reyes, C., Ye, S., Kim, M. J., & Wiley, B. J. (2017). 3d printing electronic components and circuits with conductive thermoplastic filament. *Additive Manufacturing*, *18*, 156–163.

Foldiak, P. (2003). Sparse coding in the primate cortex. *The Handbook of Brain Theory and*

Neural Networks.

Franken, T. P., Roberts, M. T., Wei, L., Golding, N. L., & Joris, P. X. (2015). In vivo coincidence detection in mammalian sound localization generates phase delays. *Nature Neuroscience*, *18*(3), 444.

Frégnac, Y., & Bathellier, B. (2015). Cortical correlates of low-level perception: from neural circuits to percepts. *Neuron*, *88*(1), 110–126.

Fu, Q.-Y., Liang, Y., Su, Y.-Y., & Wang, T. (2009). An preliminary investigation of speech-evoked abr with two stimulus intensities. In *Bioinformatics and biomedical engineering, 2009. icbbe 2009. 3rd international conference on* (pp. 1–3).

Genther, D. J., Betz, J., Pratt, S., Martin, K. R., Harris, T. B., Satterfield, S., … others (2015). Association between hearing impairment and risk of hospitalization in older adults. *Journal of the American Geriatrics Society*, *63*(6), 1146–1152.

Genther, D. J., Frick, K. D., Chen, D., Betz, J., & Lin, F. R. (2013). Association of hearing loss with hospitalization and burden of disease in older adults. *Journal of the American Medical Association*, *309*(22), 2322–2324.

Gholami-Boroujeny, S., Fallatah, A., Heffernan, B. P., & Dajani, H. R. (2016). Neural network-based adaptive noise cancellation for enhancement of speech auditory brainstem responses. *Signal, Image and Video Processing*, *10*(2), 389–395.

Giguère, C. (1998, December). *Computational modelling of OHC damage: implications for hearing aid signal processing.* Eindhoven, The Netherlands.

Giguère, C., & Smoorenburg, G. F. (1999). Computational modeling of outer hair cell damage: Implications for hearing aid signal processing. In *Psychophysics, physiology and models of hearing* (pp. 155–164). World Scientific.

Giraud, A.-L., & Poeppel, D. (2012). Speech perception from a neurophysiological perspective. In *The human auditory cortex* (pp. 225–260). Springer.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (ffr) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for in Otolaryngology*, *12*(6), 767–782.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., … Stiegemann, B. (2007). Low-frequency and high-frequency cochlear nonlinearity in humans. *The Journal of the Acoustical Society of America*, *122*(3), 1671–1680.

Gramley, V. (2010, May). *Acoustic Phonetics.* Retrieved 2016-11-25, from `http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/source-filter.jpg`

Grimsley, C. A., Sanchez, J., & Sivaramakrishnan, S. (2013). Midbrain local circuits shape sound intensity codes. *Frontiers in Neural Circuits*, *7*, 174.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752.

Hakizimana, P., Brownell, W. E., Jacob, S., & Fridberger, A. (2012). Sound-induced length changes in outer hair cell stereocilia. *Nature Communications*, *3*, 1094.

Hall, J. W. I. (2013, September). *Application of ABR in Objective Assessment of Infant Hearing.* Retrieved 2017-04-12, from `http://www.audiologyonline.com/articles/application-abr-in-objective-assessment-12079`

Hamadicharef, B. (2010). Brain-computer interface (bci) literature-a bibliometric study. In *Information sciences signal processing and their applications (isspa), 2010 10th international conference on* (pp. 626–629).

Hardcastle, W. J., Laver, J., & Gibbon, F. E. (2010). *The handbook of phonetic sciences* (Vol. 1). John Wiley & Sons.

Harrington, J., & Cassidy, S. (2012). *Techniques in speech acoustics* (Vol. 8). Springer Science & Business Media.

Hart, H. C., Hall, D. A., & Palmer, A. R. (2003). The sound-level-dependent growth in the extent of fmri activation in heschl's gyrus is different for low-and high-frequency tones. *Hearing Research*, *179*(1-2), 104–112.

Hassaan, M. R. (2011). Aided evoked cortical potential: An objective validation tool for hearing aid benefit. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, *12*(3), 155–161.

Hassaan, M. R., Ibraheem, O. A., & Galhom, D. H. (2016). Brainstem encoding of aided speech in hearing aid users with cochlear dead region (s). *International Archives of Otorhinolaryngology*, *20*(3), 226–234.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Second ed.). Springer New York.

Hatzopoulos, S., Prosser, S., Ciorba, A., Giarbini, N., Kochanek, K., Sliwa, L., … Martini, A. (2009). Threshold estimation in adult normal-and impaired-hearing subjects using auditory steady-state responses. *Medical Science Monitor*, *16*(1), CR21–CR27.

Hearing4all - Task Group 7: Brain Computer Interface for hearing devices. (2017, January). Retrieved 2017-01-10, from `http://hearing4all.eu/EN/Research/C/Task-Group-7.php`

Heffernan, B., Dajani, H., & Giguère, C. (2017). Exploring the encoding of sound level in the speech-evoked frequency following response to four synthetic english vowels. *Journal of Hearing Science*, *7*(2), 50.

Heffernan, B., Giguère, C., & Dajani, H. (2017). Towards developing a brain-computer interface for automatic hearing aid fitting based on the speech-evoked frequency following response. In *Neurotechnix 2017 - extended abstracts - neurotechnix,* (p. 3-4). SciTePress.

Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., & Picton, T. W. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topography*, *15*(2), 69–86.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218–1227.

Hoormann, J., Falkenstein, M., Hohnsbein, J., & Blanke, L. (1992). The human frequency-following response (ffr): normal variability and relation to the click-evoked brainstem response. *Hearing Research*, *59*(2), 179–188.

Hoseingholizade, S. (2015). *Investigation of the feasibilty of using chirp-evoked abr in estimation of loudness growth* (Master's dissertation). McMaster University.

Huang, H.-C., & Jansen, B. (1985). Eeg waveform analysis by means of dynamic time-warping. *International Journal of Bio-Medical Computing*, *17*(2), 135–144.

Jafarpisheh, A. S., Jafari, A. H., Abolhassani, M., Farhadi, M., Sadjedi, H., Pourbakht, A., & Shirzhiyan, Z. (2016). Nonlinear feature extraction for objective classification of complex auditory brainstem responses to diotic perceptually critical consonant-vowel syllables. *Auris Nasus Larynx*, *43*(1), 37–44.

Jalaei, B., Zakaria, M. N., Mohd Azmi, M. H. A., Nik Othman, N. A., & Sidek, D. (2017). Gender disparities in speech-evoked auditory brainstem response in healthy adults: any relation to head size? *Annals of Otology, Rhinology & Laryngology*, *126*(4), 290–295.

Johnson, K. (2011). *Acoustic and auditory phonetics*. John Wiley & Sons.

Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, *119*(11), 2623–2635.

Johnstone, B., Patuzzi, R., & Yates, G. (1986). Basilar membrane measurements and the travelling wave. *Hearing Research*, *22*(1-3), 147–153.

Joris, P. X., Carney, L. H., Smith, P. H., & Yin, T. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus. i. responses to tones at the characteristic frequency. *Journal of Neurophysiology*, *71*(3), 1022–1036.

Joris, P. X., Smith, P. H., & Yin, T. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus. ii. responses in the tuning curve tail. *Journal of Neurophysiology*, *71*(3), 1037–1051.

Kamil, R. J., & Lin, F. R. (2015). The effects of hearing impairment in older adults on communication partners: a systematic review. *Journal of the American Academy of Audiology*, *26*(2), 155–182.

Keidser, G., & Convery, E. (2016). Self-fitting hearing aids: Status quo and future predictions. *Trends in Hearing*, *20*, 1–15.

Kidmose, P., Looney, D., Jochumsen, L., & Mandic, D. (2013). Ear-eeg from generic earpieces: a feasibility study. In *Engineering in medicine and biology society (embc), 2013 35th annual international conference of the ieee* (pp. 543–546).

Kidmose, P., Looney, D., & Mandic, D. P. (2012). Auditory evoked responses from ear-eeg recordings. In *Engineering in medicine and biology society (embc), 2012 annual international conference of the ieee* (pp. 586–589).

Kobel, M., Le Prell, C. G., Liu, J., Hawks, J. W., & Bao, J. (2017). Noise-induced cochlear synaptopathy: Past findings and future studies. *Hearing Research*, *349*, 148–154.

Kochkin, S. (2010). Marketrak viii: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, *63*(1), 19–20.

Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, *339*, 40–49.

Korczak, P. A., Kurtzberg, D., & Stapells, D. R. (2005). Effects of sensorineural hearing loss and personal hearing aids on cortical event-related potential and behavioral measures of speech-sound processing. *Ear and Hearing*, *26*(2), 165–185.

Korczak, P. A., Sherlock, L. P., Hawley, M. L., & Formby, C. (2017). Relations among auditory brainstem and middle latency response measures, categorical loudness judgments, and their associated physical intensities. In *Seminars in hearing* (Vol. 38, pp. 94–114).

Korczak, P. A., Smart, J., Delgado, R., M Strobel, T., & Bradford, C. (2012). Auditory

steady-state responses. *Journal of the American Academy of Audiology*, *23*(3), 146–170.

Krachunov, S., & Casson, A. J. (2016). 3d printed dry eeg electrodes. *Sensors*, *16*(10), 1635.

Kraus, N., & Anderson, S. (2012). Hearing matters: cabr may improve hearing aid outcomes. *The Hearing Journal*, *65*(11), 56.

Kraus, N., & Anderson, S. (2014). Bigger is not better: effects of hearing loss on central processing. *The Hearing Journal*, *67*(1), 3.

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The frequency-following response: A window into human communication. In *The frequency-following response* (pp. 1–15). Springer.

Kraus, N., Anderson, S., White-Schwoch, T., Fay, R. R., & Popper, A. N. (2017). *The frequency-following response: A window into human communication* (Vol. 61). Springer.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what'and 'where'pathways in the auditory system. *Trends in Neurosciences*, *28*(4), 176–181.

Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neurotology*, *4*(2), 95–103.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research*, *166*(1-2), 192–201.

Krizman, J., Skoe, E., & Kraus, N. (2012). Sex differences in auditory subcortical function. *Clinical Neurophysiology*, *123*(3), 590–597.

Kuwada, S., Anderson, J. S., Batra, R., Fitzpatrick, D. C., Teissier, N., & D'Angelo, W. R. (2002). Sources of the scalp-recorded amplitude-modulation following response. *Journal of the American Academy of Audiology*, *13*(4), 188–204.

Langers, D. R., van Dijk, P., Schoenmaker, E. S., & Backes, W. H. (2007). fmri activation in relation to sound intensity and loudness. *Neuroimage*, *35*(2), 709–718.

Laroche, M. (2010). *A study of auditory speech processing using brainstem evoked responses under quiet and noisy conditions* (Master's dissertation). University of Ottawa (Canada).

Laroche, M., Dajani, H. R., Prévost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear and Hearing*, *34*(1), 63–74.

Léger, A. C., Reed, C. M., Desloge, J. G., Swaminathan, J., & Braida, L. D. (2015). Consonant identification in noise using hilbert-transform temporal fine-structure speech and recovered-envelope speech for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *138*(1), 389–403.

Lewis, J. D., Kopun, J., Neely, S. T., Schmid, K. K., & Gorga, M. P. (2015). Tone-burst auditory brainstem response wave v latencies in normal-hearing and hearing-impaired ears. *The Journal of the Acoustical Society of America*, *138*(5), 3210–3219.

Léger, A. C., Desloge, J. G., Braida, L. D., & Swaminathan, J. (2015). The role of recovered envelope cues in the identification of temporal-fine-structure speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *137*(1), 505–508.

Liang, R., Guo, R., Xi, J., Xie, Y., & Zhao, L. (2017). Self-fitting algorithm for digital hearing aid based on interactive evolutionary computation and expert system. *Applied*

Sciences, *7*(3), 272.

Lin, F. R. (2012). Hearing loss in older adults: who's listening? *Journal of the American Medical Association*, *307*(11), 1147–1148.

Lin, F. R., & Albert, M. (2014). Hearing loss and dementia–who is listening? *Aging Ment Health*, *18*(6), 671–673.

Lin, F. R., Thorpe, R., Gordon-Salant, S., & Ferrucci, L. (2011). Hearing loss prevalence and risk factors among older adults in the united states. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *66*(5), 582–590.

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., … others (2013). Hearing loss and cognitive decline in older adults. *Journal of the American Medical Association - Internal Medicine*, *173*(4), 293–299.

Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., & Durieux-Smith, A. (1995). Auditory steady-state responses to tones amplitude-modulated at 80–110 hz. *The Journal of the Acoustical Society of America*, *97*(5), 3051–3063.

Looney, D., Kidmose, P., & Mandic, D. P. (2014). Ear-eeg: user-centered and wearable bci. In *Brain-computer interface* (pp. 41–50). Springer.

Looney, D., Kidmose, P., Morrell, M. J., & Mandic, D. P. (2014). Ear-eeg: Continuous brain monitoring. In *Brain-computer interface* (pp. 63–71). Springer.

Lu, X., Wang, P., Niyato, D., Kim, D. I., & Han, Z. (2016). Wireless charging technologies: Fundamentals, standards, and network applications. *IEEE Communications Surveys & Tutorials*, *18*(2), 1413–1452.

Mathewson, K. E., Harrison, T. J., & Kizuk, S. A. (2017). High and dry? comparing active dry eeg electrodes to active and passive wet electrodes. *Psychophysiology*, *54*(1), 74–82.

May, B. J., Prell, G. S. L., & Sachs, M. B. (1998). Vowel representations in the ventral cochlear nucleus of the cat: effects of level, background noise, and behavioral state. *Journal of Neurophysiology*, *79*(4), 1755–1767.

McClelland, R., & McCrea, R. (1979). Intersubject variability of the auditory-evoked brain stem potentials. *Audiology*, *18*(6), 462–471.

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*(5), 360–368.

McPherson, B. (2011). Innovative technology in hearing instruments: matching needs in the developing world. *Trends in Amplification*, *15*(4), 209–214.

Mecklenburger, J., & Groth, T. (2016). Wireless technologies and hearing aid connectivity. In *Hearing aids* (pp. 131–149). Springer.

Mick, P., Kawachi, I., & Lin, F. R. (2014). The association between hearing loss and social isolation in older adults. *Otolaryngology–Head and Neck Surgery*, *150*(3), 378–384.

Middlebrooks, J. (2014). Auditory system: central pathways.

Minary, P. (2016, November). *Journey into the world of hearing.* Retrieved 2017-01-31, from `http://www.cochlea.eu/en`

Młynarski, W., & McDermott, J. H. (2018). Learning midlevel auditory codes from natural sound statistics. *Neural Computation*, *30*(3), 631–669.

Mobley, K. J., & Gibson, E. (2000, September). *Tone Burst Evoked Potentials: Clinical Applications Karen J. Mobley Edie Gibson.* Retrieved 2017-02-17, from `http://www.audiologyonline.com/articles/tone-burst-evoked-potentials-clinical-1279`

Munro, K. J., & Mueller, H. G. (2016). Clinical verification of hearing aid performance. In *Hearing aids* (pp. 253–289). Springer.

Nabelek, A. K., & Letowski, T. R. (1985). Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *Journal of Speech and Hearing Disorders*, *50*(2), 126–131.

Nelson, P. B., & VanTasell, D. (2016). Self-fitting hearing aids reveal strong individual preferences in noisy rooms. *The Journal of the Acoustical Society of America*, *140*(4), 3070–3070.

Nikiforidis, G. C., Koutsojannis, C. M., Varakis, J. N., & Goumas, P. D. (1993). Reduced variance in the latency and amplitude of the fifth wave of auditory brain stem response after normalization for head size. *Ear and Hearing*, *14*(6), 423–428.

Olson, E. S., Duifhuis, H., & Steele, C. R. (2012). Von békésy and cochlear mechanics. *Hearing Research*, *293*(1-2), 31–43.

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: a call to action. *Bulletin of the World Health Organization*, *92*, 367–373.

Palmer, A. R. (2009). *Processing in the cochlear nucleus.* Medical Council - Institute of Hearing. Retrieved from `https://health.uconn.edu/meds5377/wp-content/uploads/sites/151/2017/07/Palmer-Cochlear-nucleus-1-3.pdf`

Park, E., Cho, M., & Ki, C.-S. (2009). Correct use of repeated measures analysis of variance. *The Korean Journal of Laboratory Medicine*, *29*(1), 1–9.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184.

Phatak, S. A., Yoon, Y.-s., Gooler, D. M., & Allen, J. B. (2009). Consonant recognition loss in hearing impaired listeners. *The Journal of the Acoustical Society of America*, *126*(5), 2683–2694.

Picton, T. W. (2010). *Human auditory evoked potentials*. Plural Publishing.

Plyler, P. N., & Ananthanarayan, A. (2001). Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*, *12*(10), 523–533.

Popelka, G. R., & Moore, B. C. (2016). Future directions for hearing aid development. In *Hearing aids* (pp. 323–333). Springer.

Prévost, F., Laroche, M., Marcoux, A., & Dajani, H. (2013). Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. *Clinical Neurophysiology*, *124*(1), 52–60.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., & White, L. E. (2012). Neuroscience. *Sunderland, Mass.: Sinauer Associates*.

Quatieri, T. F. (2006). *Discrete-time speech signal processing: principles and practice*. Pearson Education India.

Rabang, C. F., Parthasarathy, A., Venkataraman, Y., Fisher, Z. L., Gardner, S. M., & Bartlett, E. L. (2012). A computational model of inferior colliculus responses to amplitude modulated sounds in young and aged rats. *Frontiers in Neural Circuits*, *6*, 77.

Rabiner, L. R., & Juang, B.-H. (1993). *Fundamentals of speech recognition* (Vol. 14). PTR

Prentice Hall Englewood Cliffs.

Rao, R. B., Fung, G., & Rosales, R. (2008). On the dangers of cross-validation. an experimental evaluation. In *Proceedings of the 2008 siam international conference on data mining* (pp. 588–596).

Rauschecker, J. P. (2015a). Auditory Cortex. In A. W. Toga (Ed.), *Brain Mapping* (pp. 299–304). Waltham: Academic Press. doi: 10.1016/B978-0-12-397025-1.00226-8

Rauschecker, J. P. (2015b). Early Auditory Processing. In A. W. Toga (Ed.), *Brain Mapping* (pp. 537–542). Waltham: Academic Press. doi: 10.1016/B978-0-12-397025-1.00042-7

Recio, A., & Rhode, W. S. (2000). Representation of vowel stimuli in the ventral cochlear nucleus of the chinchilla. *Hearing Research*, *146*(1-2), 167–184.

Reichenbach, C. S., Braiman, C., Schiff, N. D., Hudspeth, A., & Reichenbach, T. (2016). The auditory-brainstem response to continuous, non-repetitive speech is modulated by the speech envelope and reflects speech processing. *Frontiers in Computational Neuroscience*, *10*, 47.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational. *Educational Review*, *6*(2), 135–147.

Röhl, M., & Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. *Journal of the Association for Research in Otolaryngology*, *13*(3), 369–379.

Rouiller, E., De Ribaupierre, Y., Morel, A., & De Ribaupierre, F. (1983). Intensity functions of single unit responses to tone in the medial geniculate body of cat. *Hearing Research*, *11*(2), 235–247.

Sachs, M. B., Bruce, I. C., Miller, R. L., & Young, E. D. (2002). Biological basis of hearing-aid design. *Annals of Biomedical Engineering*, *30*(2), 157–168.

Sadagopan, S., & Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. *Journal of Neuroscience*, *28*(13), 3415–3426.

Sadeghian, A. (2012). *Classification of speech evoked responses to english vowels* (Masters dissertation). Carleton University.

Sadeghian, A., Dajani, H. R., & Chan, A. D. (2011). Classification of english vowels using speech evoked potentials. In *Engineering in medicine and biology society, embc, 2011 annual international conference of the ieee* (pp. 5000–5003).

Sadeghian, A., Dajani, H. R., & Chan, A. D. (2015). Classification of speech-evoked brainstem responses to english vowels. *Speech Communication*, *68*, 69–84.

Salomons, E. M., & Janssen, S. A. (2011). Practical ranges of loudness levels of various types of environmental noise, including traffic noise, aircraft noise, and industrial noise. *International Journal of Environmental and Public Health*, *8*(6), 1847–1864.

Sardari, S., Jafari, Z., Haghani, H., & Talebi, H. (2015). Hearing aid validation based on 40 hz auditory steady-state response thresholds. *Hearing Research*, *330*, 134–141.

Sayles, M., & Winter, I. M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron*, *58*(5), 789–801.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., … Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9*(4), 159–197.

Serpanos, Y. C., O'malley, H., & Gravel, J. S. (1997). The relationship between loudness intensity functions and the click-abr wave v latency. *Ear and Hearing*, *18*(5), 409–419.

Shawrav, M. M., Taus, P., Wanzenboeck, H. D., Schinnerl, M., Stöger-Pollach, M., Schwarz, S., … Bertagnolli, E. (2016). Highly conductive and pure gold nanostructures grown by electron beam induced deposition. *Scientific Reports*, *6*, 34003.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302.

Smith, A. W. (2008). Demographics of hearing loss in developing countries. *Audiology in Developing Countries*, 21–50.

Solanki, J. D., Mehta, H. B., et al. (2015). Sex as a source of variance affecting auditory evoked potential. *The Egyptian Journal of Otolaryngology*, *31*(2), 111.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: neural correlates. *Journal of Cognitive Neuroscience*, *23*(9), 2268–2279.

Stapells, D., Gravel, J., & Martin, B. (1996). Thresholds for auditory brain stem responses to tones in notched noise from infants and young children with normal hearing or sensorineural hearing lo. *International Journal of Pediatric Otorhinolaryngology*, *3*(34), 289.

Stapells, D. R. (2000). Threshold estimation by the tone-evoked auditory brainstem response: a literature meta-analysis. *Journal of Speech Language Pathology and Audiology*, *24*(2), 74–83.

Strickland, E. (2016, July). *In-Ear EEG Makes Unobtrusive Brain-Hacking Gadgets a Real Possibility.* Retrieved 2016-11-22, from `http://spectrum.ieee.org/the-human-os/biomedical/devices/in-ear-eeg-makes-unobtrusive-brain-hacking-gadgets-a-real-possibility`

Svirsky, M. A. (2000). Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues. *The Journal of the Acoustical Society of America*, *107*(3), 1521–1529.

Syka, J. (2010). Subcortical responses to species-specific vocalizations. In *Handbook of behavioral neuroscience* (Vol. 19, pp. 99–112). Elsevier.

Takagi, H., & Ohsaki, M. (1999). Iec-based hearing aid fitting. In *Systems, man, and cybernetics, 1999. ieee smc'99 conference proceedings. 1999 ieee international conference on* (Vol. 3, pp. 657–662).

Takagi, H., & Ohsaki, M. (2007). Interactive evolutionary computation-based hearing aid fitting. *IEEE Transactions on Evolutionary Computation*, *11*(3), 414–427.

Tanaka, T. (2016). Information maximization in a feedforward network replicates the stimulus preference of the medial geniculate and the auditory cortex. In *International conference on neural information processing* (pp. 183–190).

Tanaka, T., & Nakamura, K. (2013). Information maximization principle explains the emergence of complex cell-like neurons. *Frontiers in Computational Neuroscience*, *7*, 165.

Teng, X., Tian, X., & Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Scientific Reports*, *6*, 34390.

Thwaites, A., Glasberg, B. R., Nimmo-Smith, I., Marslen-Wilson, W. D., & Moore, B. C. (2016). Representation of instantaneous and short-term loudness in the human cortex. *Frontiers in Neuroscience*, *10*, 183.

Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hearing Research*, *348*, 1–15.

Tremblay, K. L. (2015). The ear–brain connection: Older ears and older brains. *American Journal of Audiology*, *24*(2), 117–120.

Tremblay, K. L., Kalstein, L., Billings, C. J., & Souza, P. E. (2006). The neural representation of consonant-vowel transitions in adults who wear hearing aids. *Trends in Amplification*, *10*(3), 155–162.

Turnip, A., Hutagalung, S. S., Pardede, J., & Soetraprawata, D. (2013). P300 detection using a multilayer neural network classifier based on adaptive feature extraction. *International Journal of Brain and Cognitive Sciences*, *2*(5), 63–75.

Ulfendahl, M., & Flock, A. (1998). Outer hair cells provide active tuning in the organ of corti. *Physiology*, *13*(3), 107–111.

Uppenkamp, S., & Röhl, M. (2014). Human auditory neuroimaging of intensity and loudness. *Hearing Research*, *307*, 65–73.

Uykan, Z., & Koivo, H. N. (2000). Unsupervised learning of sigmoid perceptron. In *Acoustics, speech, and signal processing, 2000. icassp'00. proceedings. 2000 ieee international conference on* (Vol. 6, pp. 3486–3489).

Valentin, O., Viallet, G., & Voix, J. (2017). Eartrodes: Towards a wireless in-ear custom-fitted brain computer interface. *Canadian Acoustics= Acoustique Canadienne*, *45*(3), 142–143.

Van Dun, B., Kania, A., & Dillon, H. (2016). Cortical auditory evoked potentials in (un) aided normal-hearing and hearing-impaired adults. In *Seminars in hearing* (Vol. 37, p. 9).

Van Eeckhoutte, M., Wouters, J., & Francart, T. (2016). Auditory steady-state responses as neural correlates of loudness growth. *Hearing Research*, *342*, 58–68.

VanRullen, R., & Dubois, J. (2011). The psychophysics of brain rhythms. *Frontiers in Psychology*, *2*, 203.

Van Santen, J. P. (1992). Contextual effects on vowel duration. *Speech Communication*, *11*(6), 513–546.

Vázquez, Y., Salinas, E., & Romo, R. (2013). Transformation of the neural code for tactile detection from thalamus to cortex. *Proceedings of the National Academy of Sciences*, *110*(28), E2635–E2644.

Vidya, R., Nasira, G., & Priyankka, R. J. (2014). Sparse coding: a deep learning using unlabeled data for high-level representation. In *Computing and communication technologies (wccct), 2014 world congress on* (pp. 124–127).

Von Békésy, G. (1963). Hearing theories and complex sounds. *The Journal of the Acoustical Society of America*, *35*(4), 588–601.

Wang, C., & Madou, M. (2005). From mems to nems with carbon. *Biosensors and Bioelectronics*, *20*(10), 2181–2187.

Wang, D. (2017). Deep learning reinvents the hearing aid. *IEEE Spectrum*, *54*(3), 32–37.

Wang, D., Zhou, Q., & Hussain, A. (2016). Deep and sparse learning in speech and language processing: An overview. In *International conference on brain inspired cognitive systems* (pp. 171–183).

Wang, X., Lu, T., Bendor, D., & Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience*, *157*(2), 484–493.

Williams-Sanchez, V. A. (2014). *Word recognition in noise among young and older listeners: A combined behavioral and electrophysiological study* (Doctoral dissertation). University of South Florida.

Wilson, K. G., & Stelmack, R. M. (1982). Power functions of loudness magnitude estimations and auditory brainstem evoked responses. *Perception & Psychophysics*, *31*(6), 561–565.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*(1), 1–11.

World Health Organization, et al. (2015). Hearing loss due to recreational exposure to loud sounds: a review.

Xu, Z.-m., Cheng, W.-x., & Yao, Z.-h. (2014). Prediction of frequency-specific hearing threshold using chirp auditory brainstem response in infants with hearing losses. *International Journal of Pediatric Otorhinolaryngology*, *78*(5), 812–816.

Yi, H. G., Xie, Z., Reetzke, R., Dimakis, A. G., & Chandrasekaran, B. (2017). Vowel decoding from single-trial speech-evoked electrophysiological responses: A feature-based machine learning approach. *Brain and Behavior*, *7*(6).

Yost, W. (2013). *Fundamentals of Hearing: An Introduction* (5th ed. edition ed.). Brill Academic Pub.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *66*(5), 1381–1403.

Zaitoun, M., Purcell, A., & Cumming, S. (2015). Investigation of the existing testing practices in the auditory brain-stem response (abr): An online survey. *Journal of International Advanced Otology*, *11*.

Zenker Castro, F., & Barajas de Prat, J. J. (2008). The Role of Auditory Steady-State Responses in Fitting Hearing Aids. In *Auditory Steady-State Responses and Hearing Device Fitting.* Plural Publishing Inc.

Zenker Castro, F., Juan Barajas de Prat, J., & Larumbe Zabala, E. (2008). Loudness and auditory steady-state responses in normal-hearing subjects. *International Journal of Audiology*, *47*(5), 269–275.