

**TOWSON UNIVERSITY
OFFICE OF GRADUATE STUDIES**

**THE EFFECTS OF ADVERSE LISTENING CONDITIONS ON THE
SUBCORTICAL NEURAL ENCODING OF SPEECH STIMULI IN
NORMAL-HEARING ADULTS**

by

Randi L. Cropper

**A thesis
Presented to the faculty of
Towson University
in partial fulfillment
of the requirements for the degree
Doctor of Audiology**

Department of Audiology, Speech-Language Pathology and Deaf Studies

**Towson University
Towson, Maryland 21252**

May 2016

TOWSON UNIVERSITY
OFFICE OF GRADUATE STUDIES

THESIS APPROVAL PAGE

This is to certify that the thesis prepared by Randi Cropper B.S., Au.D. Candidate, entitled The Effect of Adverse Listening Conditions on Subcortical Neural Encoding of Speech in Normal Hearing Individuals has been approved by the thesis committee as satisfactorily completing the thesis requirements for the degree Doctor of Audiology (Au.D.).

A. Sarada

Saradha Ananthakrishnan, Ph.D.
Chair, Thesis Committee

5-14-2015

Date

PEGGY KORCZAK [signed for Dr. Korczak]
[- Saradha Ananthakrishnan]

Peggy Korczak, Ph.D.
Committee Member

5-14-2015

Date

Stephen Pallett

Stephen Pallett, Au.D.
Committee Member

5/13/2015

Date

Janet V. Delany

Janet Delany
Dean of Graduate Studies

5-15-15

Date

ACKNOWLEDGEMENTS

I would like to thank my thesis committee for their endless support and motivation throughout this process. Especially Dr. Ananthakrishnan, who met with me individually to help me understand this complex material as well as organize my thoughts, and Dr. Evitts who made time to review basic speech acoustics with us so that we could have a better understanding of the material. I would like to thank my family and friends for providing me with the encouragement to get through this tough process. I would like to thank my cohort who kept me laughing and helped to relieve some of the stress throughout this program and thesis process. Specifically my roommate, Tessa, who was always there to keep me grounded and motivate me to get through difficult times. My two group members, Laura and Donnie who were there to brainstorm and talk through this subject matter. Lastly I would like to thank my fiancée, Ryan, for sitting with me through countless hours of writing, re-writing and stressing over this project. I love you all and I cannot thank you enough for your support.

ABSTRACT

The Effect of Adverse Listening Conditions on the Sub-Cortical Neural Encoding of Speech Stimuli in Normal Hearing Adults

Randi Cropper

Objectives

To determine the difference in the effects of background noise and reverberation on the sub-cortical neural encoding of the speech stimulus /u/ using the Frequency Following Response (FFR). Energy related to both the fundamental frequency as well as first formant of the stimulus preserved in the FFR was measured in order to better understand the breakdown of speech in adverse listening conditions.

Design

The FFR was recorded to 6 normal hearing adults (aged 24-25 years) in response to the vowel stimulus /u/. Each subject underwent two test sessions. The first session recorded the response to the stimulus in the presence of three levels of reverberation as well as a quiet condition involving no reverberation. The second session recorded the response to the stimulus in the presence of three levels of background noise as well as a quiet or no noise condition. Temporal waveforms, FFTs, and individual amplitude data for both F0 and F1 were generated for each test condition.

Results

As expected, as the severity of the condition worsened, the response energy at the F0 decreased. This was seen for both the background noise and reverberation test conditions. In contrast, there were some differences in F1 encoding that occurred as a function of type of adverse listening condition. As expected, the energy at the F1 decreased as background noise condition worsened. However, the energy at the F1 *increased* as the reverberation condition worsened. This was an unexpected finding. The variability in the data, as reflected in the standard deviation values, was fairly consistent across all test conditions except for F1 data of reverberation. This change in variability could have played a role in the unexpected finding for that condition.

Conclusion

The results of the current study suggest that degraded neural encoding abilities at the F₀ and first formant may play a role in the speech perception difficulties individuals with sensorineural hearing loss experience.

TABLE OF CONTENTS

I.	LIST OF TABLES.....	ix
II.	LIST OF FIGURES.....	x
III.	CHAPTER 1: Introduction.....	1
IV.	CHAPTER 2: Literature Review.....	3
	Auditory Evoked Potentials.....	3
	Classification of AEPs.....	3
	The Fundamentals of Speech Acoustics.....	5
	Neural Encoding.....	9
	Firing Patterns of 8 th Nerve Fibers.....	13
	What is the FFR?	14
	History of the FFR.....	15
	Neural Generators of the FFR.....	19
	Recent Research Utilizing the FFR.....	25
	Recording and Measuring the FFR to Envelope and TFS Cues.....	26
	Technical Parameters.....	31
	Stimulus Parameters.....	32
	Stimulus type.....	32
	Stimulus intensity.....	35
	Stimulus polarity.....	38
	Stimulus rate.....	40

Recording Parameters.....	42
Electrode montage.....	42
Analog band-pass filter settings.....	45
Sampling rate.....	46
Number of sweeps.....	47
Length of the averaging window.....	48
Artifact.....	48
Subject Parameters.....	50
Attention.....	50
Sleep state.....	52
Age/maturation effects.....	52
Psychoacoustics: reverberation and background noise.....	53
Background Noise.....	54
Effects of background noise on speech acoustics.....	54
Effects of background noise on speech understanding.....	57
Effects of background noise on neural encoding.....	62
Reverberation.....	64
Effects of reverberation on speech acoustics.....	64
Effects of reverberation on speech understanding.....	67
Effects of reverberation on brainstem FFR/neural encoding of speech.....	68
Differences in the Effects of Background Noise and Reverberation on Speech Acoustics and Perception.....	69

	Objectives of the current study.....	70
V.	CHAPTER 3: Methods.....	72
	Participants.....	72
	Stimulus.....	72
	Procedures.....	73
	FFR Data Analysis.....	74
VI.	CHAPTER 4: Results.....	77
	FFR _{ENV} Responses Related to the Fundamental Frequency (F0).....	77
	Temporal waveforms- background noise condition.....	77
	Temporal waveforms- reverberation condition.....	79
	Spectral analyses- background noise condition.....	81
	Spectral analysis- reverberation condition.....	83
	Mean and Individual Amplitude Data at the Fundamental Frequency (F0).....	84
	Background noise condition—amplitude data trends.....	86
	Reverberation condition—amplitude data trends.....	87
	FFR _{TFS} Responses Related to the First Formant (F1).....	89
	Temporal waveforms- background noise condition.....	89
	Temporal waveforms- reverberation condition.....	90
	Spectral analysis- background noise.....	92
	Spectral analysis- reverberation condition.....	94
	Mean and Individual Amplitude Data at the First Formant (F1).....	95
	Background noise condition—SNR data trends.....	98

	Reverberation condition—SNR data trends.....	99
VII.	CHAPTER 5: Discussion.....	102
	Neural Encoding of Complex Auditory Speech Signals.....	102
	Brainstem Neural Encoding.....	104
	FFRENV.....	104
	FFRTFS.....	105
	Effects of BGN & Reverberation on FFRENV.....	107
	Background noise and envelope encoding.....	107
	Reverberation and envelope encoding.....	109
	Effects of BGN & Reverberation on FFRTFS.....	111
	Background noise and TFS encoding.....	111
	Reverberation and TFS encoding.....	113
	Clinical Implications.....	114
	Limitations.....	115
	Future Research.....	115
VIII.	APPENDICES.....	117
	APPENDIX A: Informed Consent Form.....	118
	APPENDIX B: IRB Approval Form.....	119
IX.	REFERENCES.....	121
X.	CURRICULUM VITA.....	131

LIST OF TABLES

Table 1. <i>A comparison of FFR latencies per electrode site in cats and humans (from Smith et al., 1975)</i>	24
Table 2. <i>Differences between the effects of background noise and reverberation on the speech signal</i>	70
Table 3. <i>Individual subjects' amplitude values at the fundamental frequency (F_0) for both background noise and reverberation conditions</i>	85
Table 4. <i>Individual subjects' derived SNR values at the first formant (F_1) for both background noise and reverberation conditions</i>	97
Table 5. <i>Summary of the results for both adverse listening conditions</i>	101

LIST OF FIGURES

<i>Figure 1.</i> Harmonic structure of the vowel /u/.....	6
<i>Figure 2.</i> Formant structure of a vowel.....	8
<i>Figure 3.</i> Band-pass filters found along the basilar membrane in reversed order from low to high frequency.....	10
<i>Figure 4.</i> In-coming simple stimulus (250 Hz pure tone).....	10
<i>Figure 5.</i> A complex stimulus as it is separated and sent through their respective bins.....	12
<i>Figure 6.</i> Envelope (E) as it modulates the complex temporal fine structure (TFS) waveform.....	12
<i>Figure 7.</i> Phase-locking response pattern to a 250 Hz stimulus.....	13
<i>Figure 8.</i> Monaural tracings from the round window and cochlear nucleus for both the left and right side (Marsh & Worden, 1968).....	16
<i>Figure 9.</i> Amplitudes of the FFR as a function of frequency and intensity of the stimulus (Marsh & Worden, 1968).....	17
<i>Figure 10.</i> FFR response waveforms to different frequency tone burst stimuli (Moushegian et al., 1973).....	19
<i>Figure 11.</i> Stimulus intensity as it affects the FFR in cat from Smith et al. (197.....	23
<i>Figure 12.</i> Recordings before, during and after cooling of the ICs at different electrode site (Smith et al., 1975).....	25

<i>Figure 13.</i> Spectrogram of the vowel /u/.....	27
<i>Figure 14.</i> Rarefaction trace.....	28
<i>Figure 15.</i> Condensation trace.....	28
<i>Figure 16.</i> Added waveform rarefaction + condensation.....	29
<i>Figure 17.</i> FFT of summed waveform.....	30
<i>Figure 18.</i> Subtraction waveform condensation – rarefaction.....	31
<i>Figure 19.</i> FFT of subtracted waveform.....	31
<i>Figure 20.</i> Diagram of simple vs complex stimuli.....	32
<i>Figure 21.</i> Time-varying stimulus.....	33
<i>Figure 22.</i> Sustained stimulus.....	34
<i>Figure 23.</i> FFRs recorded to different stimulus intensities (Moushegian et al., 1973).....	36
<i>Figure 24.</i> FFRs recorded to several different stimulus intensities (Krishnan, 2002).....	37
<i>Figure 25.</i> FFR recorded to several different stimulus intensities (Akhoun et al., 2008).....	25
<i>Figure 26.</i> Vertical montage.....	43
<i>Figure 27.</i> Horizontal montage.....	44
<i>Figure 28.</i> FFTs of vowel in quiet (left panel) vs vowel in +6 dB SNR (right panel).....	56
<i>Figure 29.</i> From Assmann & Summerfield (2003).....	57
<i>Figure 30.</i> FFTs of two separate vowel stimuli.....	59

<i>Figure 31.</i> FFTs of a vowel in BGN (top panel) and the vowel it is perceived as (bottom panel).....	60
<i>Figure 32.</i> From (Assmann & Summerfield, 2003).....	66
<i>Figure 33.</i> Grand averaged temporal waveforms for varying degrees of background noise.....	78
<i>Figure 34.</i> Grand averaged temporal waveforms for varying degrees of reverberation.....	80
<i>Figure 35.</i> Grand averaged FFTs for varying degrees of background noise.....	82
<i>Figure 36.</i> Grand averaged FFTs for varying degrees of reverberation	83
<i>Figure 37.</i> Mean amplitude values for the F_0 across four conditions of background noise.....	87
<i>Figure 38.</i> Mean amplitude values for the F_0 across four conditions of reverberation.....	88
<i>Figure 39.</i> Grand averaged temporal waveforms for varying degrees of background noise.....	89
<i>Figure 40.</i> Grand averaged temporal waveforms for varying degrees of reverberation.....	91
<i>Figure 41.</i> Grand averaged FFTs for varying degrees of background Noise.....	93
<i>Figure 42.</i> Grand averaged FFTs for varying degrees of reverberation	94
<i>Figure 43.</i> Mean SNR values for the F_1 response across four background noise conditions.....	98
<i>Figure 44.</i> Mean SNR values for the F_1 response across four reverberation	

conditions.....100

Figure 45. Demonstration of a complex stimulus and its resulting outputs as it passes through various auditory filters along the basilar membrane..... 104

Figure 46. Demonstrates a complex waveform as it passes through the auditory filters along the basilar membrane along with the addition of background noise.....109

Chapter 1:

Introduction

A common complaint of many individuals with hearing loss is difficulty understanding speech in the presence of adverse listening conditions such as background noise and/or reverberation. The combination of a poor acoustic environment and several competing sounds can make understanding speech in a complex environment difficult for even a normal hearing person. This is thought to be due to the degradation of the speech signal that occurs in the presence of such environments (Assman & Summerfield, 2003). For the hearing impaired, amplification devices do little to help in these environments. Although technology is continuously improving, it has yet to acquire the ability to filter between wanted and unwanted signals. In order to advance amplification technology further, it is necessary to first understand where the breakdown of speech understanding occurs under difficult listening conditions.

One way we can measure the encoding of speech at the sub-cortical level is by use of an electrophysiologic response known as the Frequency Following Response, or FFR. The FFR reflects the activity of phase-locking neurons at the level of the rostral brainstem, which encode both simple and complex stimuli. Specifically, the FFR provides information directly related to two important aspects of the speech signal: 1. the envelope and 2. the temporal fine structure.

In recent years, research has focused on recording the FFR in response to speech stimuli (Banai, Hornickel, Skoe, Nicol, Zecker... & Kraus, 2009; Cunningham, Nicol, Zecker, Bradlow, & Kraus, N., 2001; Johnson, Nicol, & Kraus, 2005; Kraus & Nicol,

2005; Lorenzi, Gilbert, Carn, Garnier & Moore, 2006; Plyler & Ananthanarayan, 2001; Russo, Nicol, Musacchia, & Kraus, 2004). These studies, however, have mainly focused on the response to speech that is presented in a quiet listening environment at a high intensity level. Behavioral studies investigating the understanding of speech in the presence of complex listening environments have found strong correlations between degraded speech environments and poor speech understanding. However, few researchers have examined what happens to this speech signal when it is processed at the sub-cortical level.

The current study is aimed at evaluating this sub-cortical encoding process after the signal has been altered by background noise and reverberation. Knowledge of these effects may provide hearing scientists with a better understanding of speech understanding errors that occur in these types of environments. In the long term, it is hoped that a better understanding of brainstem encoding of noise and reverberation-induced distortions can lead to better signal-processing strategies in amplification devices.

Prior to recording and interpreting the FFR, it is essential that an audiologist understand several topics related to this response. These topics include: basic speech acoustics; neural encoding at the level of the basilar membrane and the brainstem; history and neural generators of the response; ideal stimulus, recording and subject parameters for recording of the FFR; and predicted effects of background noise and reverberation on the FFR. The following literature review will focus on the understanding on these topics.

Chapter 2:

Literature Review

Auditory Evoked Potentials

When the brain is stimulated by an auditory stimulus, changes in its ongoing resting electrical activity, known as the electroencephalography (EEG), occur in response to this type of stimulation (Hall, 2007). These electrical changes occur as a function of time and are dependent upon the acoustic properties of the stimulus presented. Such changes in the EEG that occur in response to auditory stimuli are called Auditory Evoked Potentials (AEPs).

Classification of AEPs

There are four major classification schemes of AEPs (Picton, 1990). Specifically, AEPs can be classified based on: 1) response latency (or when the response occurs); 2) the temporal relationship of the response to the stimulus; 3) the neural generators responsible for the response; and 4) whether the response is a sensory potential or a processing contingent potential (Burkard, Don & Eggermont, 2007; Hall, 2007). The early classification of AEPs was by latency, where AEPs were categorized as short, middle and late potentials. AEPs that are recorded in 10 ms or less are considered to be short latency responses. AEPs that fall within the early/short latency category include the cochlear microphonic (CM) (< 0.5 ms), summing potential (< 1 ms), and auditory brainstem response (ABR) Wave I (< 2 ms). AEPs that fall within 10 to 50 ms are considered middle latency potentials. The Middle Latency Response (MLR) is an

example of an AEP that fits into the middle latency category. For the MLR, the latencies of waves Na, Pa and Nb occur between 10 to 50 ms after the onset of the stimulus. All AEPs that have latencies greater than 50 ms are considered to be long latency responses. An example of a long latency response is wave N1 of the slow cortical response, which occurs approximately 80-100 ms post stimulus onset.

AEPs can also be classified by the temporal relationship of the response to the stimulus; they can be described as transient, sustained, or steady-state responses (Burkard et al., 2007). Transient responses are evoked potentials elicited from the onset or offset of the stimulus. An example of a transient AEP is the click-evoked ABR. In contrast, sustained responses are evoked potentials that are elicited throughout the duration of a stimulus; i.e. the response can be seen as long as the stimulus is being presented. An example of a sustained response would be the Frequency-Following Response (FFR) (Burkard et al., 2007). Lastly, auditory steady-state potentials are evoked by rapidly repeating stimuli, with the neural response overlapping in the post-stimulus analysis window. An example of a steady-state response is the Auditory Steady-State Response (ASSR) (Burkard et al., 2007).

The third classification scheme of AEPs is related to the neural generators responsible for the response and, indirectly, the response latency. The earliest responses, such as the cochlear microphonic (CM), originate from the cochlea. The short latency responses, such as the ABR, originate from the 8th cranial nerve as well as multiple generators within the auditory brainstem area. Lastly, the middle latency and later cortical AEPs are primarily generated by the auditory cortex and its association areas, as well as certain sub-cortical structures.

The final classification scheme for AEPs is based on whether the potential is a sensory potential or a processing contingent potential (PCP) (Burkard et al., 2007; Hall, 2007). Sensory AEPs, or exogenous potentials, are obligatory responses that depend on the presence of a stimulus and are sensitive to changes in the physical properties of the stimulus. The ABR is an example of a sensory AEP, as its response amplitudes and latencies are influenced by stimulus intensity. Specifically, the response amplitude decreases and latency increases with a decrease in stimulus intensity. In contrast to sensory AEPs, processing-contingent, or endogenous potentials are those in which processing goes on beyond the obligatory sensory stages. An example of a PCP is wave P3b or the P300, which reflects not only the detection of the auditory signals but also the active discrimination of the acoustic difference between the stimuli.

Based on the classification systems mentioned above, the Frequency Following Response (FFR) is a fast latency, sustained, sensory AEP. The FFR can be used to investigate subcortical neural processing of complex signals such as speech. Audiologists wishing to successfully record and analyze the FFR must first have a basic understanding of speech acoustics, neural encoding of complex auditory signals, and neural firing patterns. These concepts will be discussed in the next section of this literature review.

The Fundamentals of Speech Acoustics

Speech is a complex waveform consisting of several different frequency components, including the fundamental frequency, harmonics, and formant frequencies. These frequency components can be visualized on a spectrogram, which is a graph plotting energy contained at frequencies as a function of time. An example of a

spectrogram plot is seen in Figure 1 below. The fundamental frequency, or F_0 , is the lowest frequency present in a complex signal (Borden & Harris, 1984). The “source-filter” theory of speech production can be used to further describe the fundamental frequency of a speech signal. Based on the “source-filter” theory of speech production, the larynx serves as the “source” of sound energy when the vocal folds come together and move apart, alternately blocking and allowing air flow from the lungs. The fundamental frequency of a speaker’s voice is determined by the rate at which the vocal folds open and close, and is unique to every speaker.

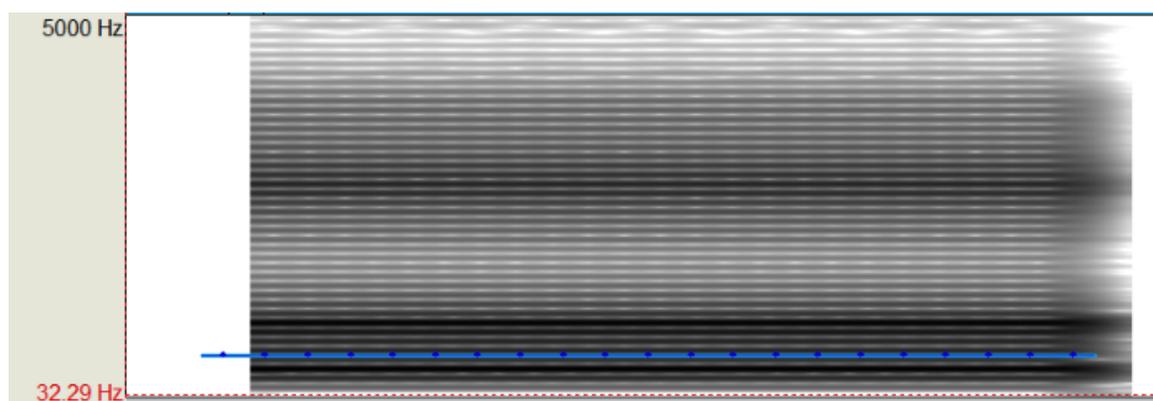


Figure 1. Harmonic structure of the vowel /u/.

In addition to the fundamental frequency, complex signals such as speech also contain energy at whole number multiples of the fundamental frequency, known as harmonics. In the spectrogram seen above, the fundamental frequency (F_0) of the vowel is 120 Hz, represented by the blue bar. In contrast, the dark grey bands that occur at whole number multiples of the fundamental frequency represent harmonics. In the complex stimulus depicted above, the harmonics can be seen at whole number multiples of 120 Hz (F_0): 240, 360, 480 Hz etc.

Apart from the fundamental frequency and harmonics, complex speech stimuli are also characterized by formant frequencies. Formant frequencies are acoustic resonances

(enhanced energy at certain frequencies) generated by the vocal tract during the production of speech. Based on the “source-filter” theory of speech production, once the airflow from the lungs passes through the vocal cords, it is filtered differentially through various articulators (i.e., the tongue, lips, throat, and teeth) depending on the speech sound being produced. Different articulators are used in the production of different sounds. For instance, the lips are brought together and the velopharyngeal flap is closed during the production of the bilabial sound /p/. In contrast, the lips remain open while the tongue tip touches the back of the upper teeth in the production of the dental sound /t/. Each time a specific sound is produced, a different set of articulators is engaged, changing the shape and thereby the resonance properties of the supralaryngeal vocal tract. The acoustic resonances of the vocal tract are unique for every speech sound and account for why listeners are able to distinguish different speech sounds produced by a single speaker. In other words, the fundamental frequency of a speaker, or the rate at which his or her vocal chords vibrate, is constant for all speech sounds. However, due to the filter action of the vocal tract, listeners will be able to distinguish the speaker’s production of the vowel /a/ (open back vowel), which has a different set of formant frequencies from the vowel /u/ (front rounded vowel). On a spectrogram, formant frequencies are typically labeled as F_1 , F_2 , F_3 and so on, as seen in Figure 2 below (Borden & Harris, 1984).

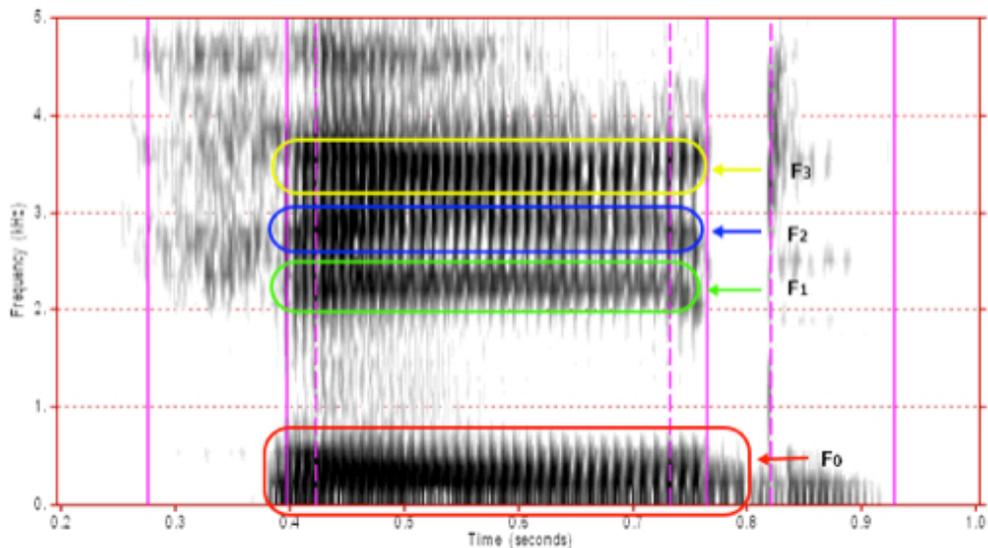


Figure 2. Formant structure of a vowel.

In the spectrogram pictured in Figure 2, the F_0 is observed at approximately 500 Hz, and F_1 , F_2 , and F_3 can be seen at 2500, 2750, and 3500 Hz, respectively. As opposed to harmonics that occur at whole number multiples of the fundamental frequency, formants occur where the greatest amount of energy is present in the signal, indicated on the spectrograph by darker bands.

As mentioned earlier, the Frequency Following Response (FFR), which can be recorded to speech stimuli, is capable of reflecting both the fundamental frequency as well as the formant structure contained in the speech signal. However, a clear understanding of the neural activity underlying the FFR is required before further discussion of this versatile subcortical response. Now that we have a good understanding of basic speech acoustics, we now need to understand the fundamentals of neural encoding of complex auditory signals.

Neural Encoding

To introduce this topic, we will first discuss neural encoding at the level of the basilar membrane. When an auditory signal arrives at the basilar membrane, it encounters a series of band-pass filters. These filters may be described in terms of their center frequency (central frequency between low and high cut-off frequencies) and filter width. The auditory filters, or “bins”, also vary in terms of their filter width. Band-pass filters are narrower near the apical end of the cochlea in the lower frequency region and grow wider towards the basal end of the cochlea in the higher frequency region. Where the filters are narrower, there is a smaller difference between the low-pass and high-pass frequency cut-offs associated with that “bin”. For example, towards the apical end, a filter could exist that contains a low-pass cut-off of 100 Hz and a high-pass cut-off of 200 Hz. Therefore, any signal that was sent to the level of the basilar membrane between 100 and 200 Hz would be sent through this “bin”. In contrast, as the filters widen, the difference between the low-pass and high-pass frequencies becomes increasingly larger (see Figure 3). For example, a filter could exist with a low-pass cut-off set to 4000 Hz and a high-pass cut-off set to 6000 Hz. In comparison to the narrower low frequency “bins”, there is a greater range of frequencies (any signal between 4000 and 6000 Hz) that could be sent through this “bin”.



Figure 3. Band-pass filters found along the basilar membrane in reversed order from low to high frequency. As the frequencies increase, the associated band-pass filters get increasingly wider. LP= Low-pass cut-off; HP= High-pass cut-off.

When a simple stimulus (i.e., a pure tone) reaches the level of the basilar membrane to be processed, it passes through its respective filter. For example, if a pure-tone stimulus of 125 Hz is presented to the basilar membrane, it would pass through the band-pass filter previously mentioned with the low and high-pass cut-offs set at 100 and 200 Hz, respectively. The filter then yields an output of a sine wave representing the simple acoustic stimulus (Figure 4).

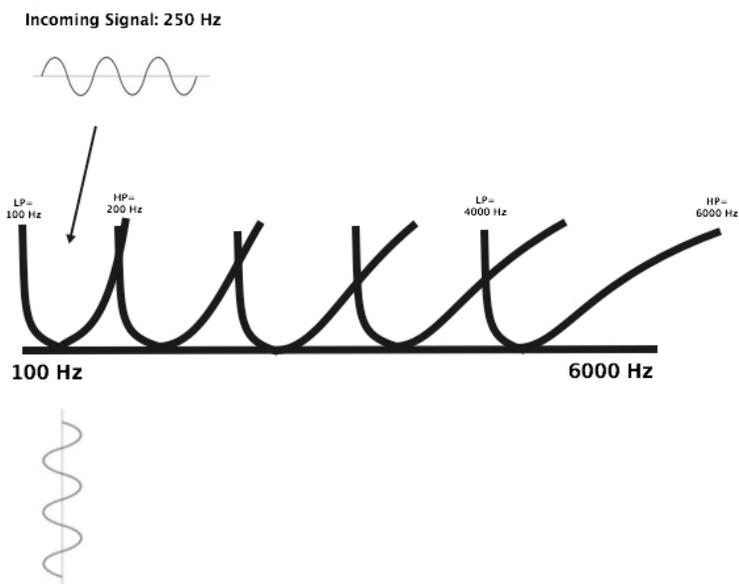


Figure 4. In-coming simple stimulus (250 Hz pure tone).

When a complex stimulus consisting of multiple frequencies reaches the level of the basilar membrane, it is first divided into its component frequencies. Each component frequency is then sent to the band-pass filter with the appropriate corresponding center frequency. That is, the low frequency components are sent to filters with low center frequencies, mid frequency components to the filters with mid center frequencies, and high frequency components to filters with high center frequencies. Due to the narrow width of low frequency filters, a single low frequency component would pass through one, narrow, low frequency filter, resulting in what is known as a “resolved” filter output consisting of a single pure tone. On the other hand, given that high frequency filters are wider, *multiple* higher frequency components will pass through the *same* high frequency auditory filter. This results in an “unresolved” filter output consisting of a complex waveform. Figure 5 below depicts what would occur in the case of a complex signal with frequencies ranging from 125- 5300 Hz, specifically, how the signal would be broken apart by frequency and filtered through separate “bins” along the basilar membrane. In this case, the 125 Hz low frequency component would go through a low frequency “bin” with a center frequency around 125 Hz. The filter output would be a pure tone at 125 Hz. Meanwhile, all of the high frequency components between 4000 and 5300 Hz would be filtered through the *same* higher frequency “bin”, resulting in a complex waveform output.

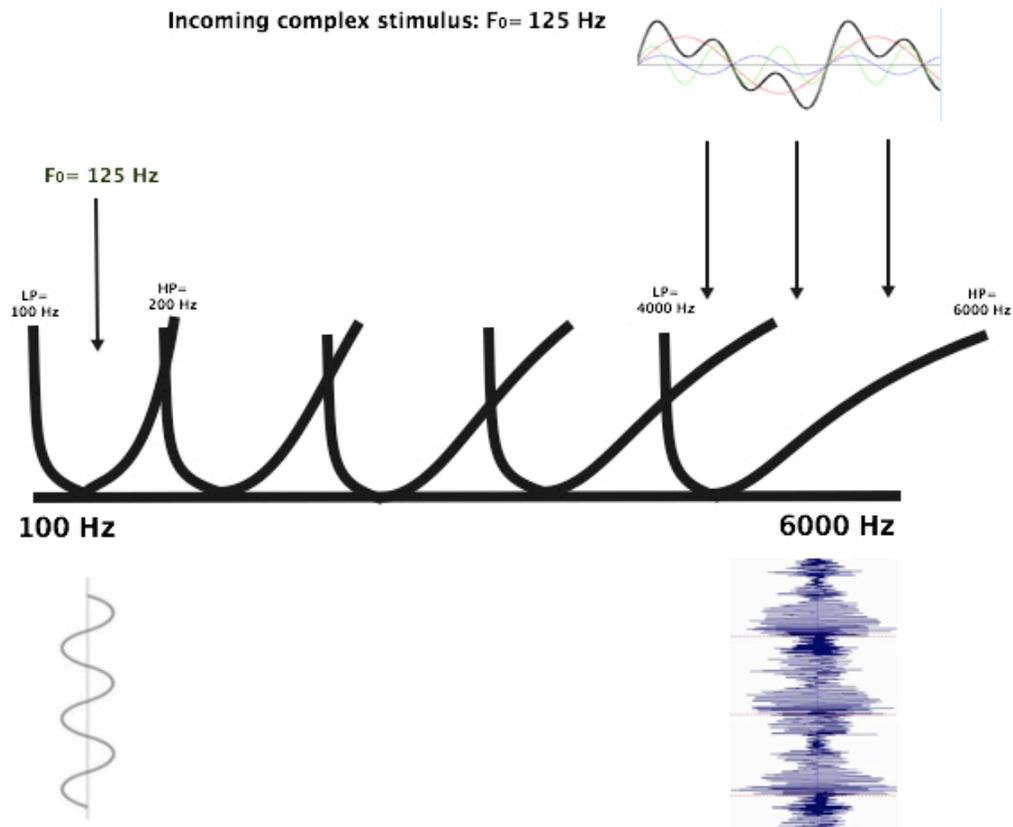


Figure 5. A complex stimulus as it is separated and sent through their respective bins.

This complex waveform output is composed of a slow-varying envelope superimposed on a rapidly-varying fine structure, as depicted in Figure 6 below.

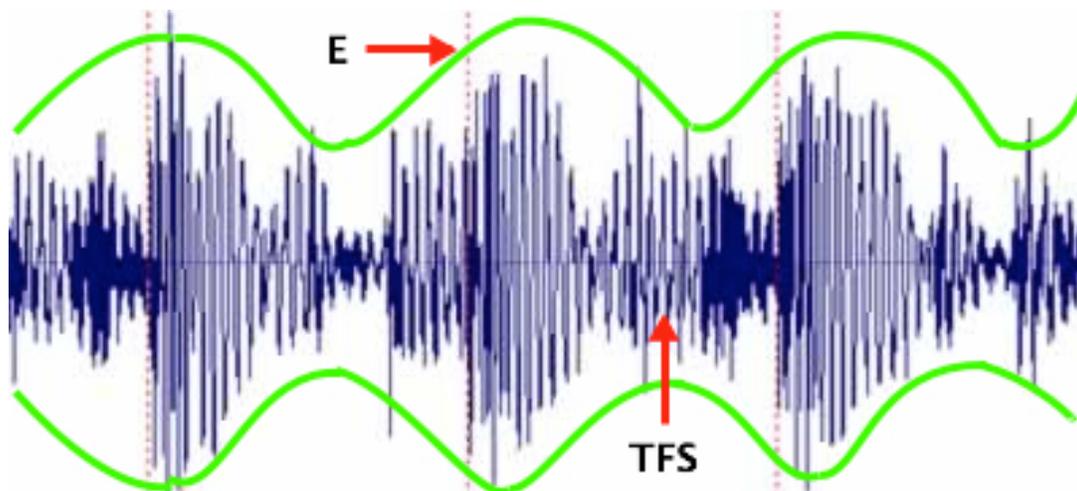


Figure 6. Envelope (E) as it modulates the complex temporal fine structure (TFS) waveform.

This temporal fine structure (TFS; representing the high frequency content of the stimulus such as the harmonics and frequency formants) is modulated by the overall rate of the envelope (representing the low frequency fundamental of the stimulus) (Moore, 2008). Both envelope and TFS cues are encoded by neural phase-locking throughout the auditory system. The phenomenon of phase-locking will be outlined in the next section of this paper.

Firing Patterns of 8th Nerve Fibers

A substantial portion of frequency encoding in the auditory system can be attributed to neural phase-locking. Phase-locking is the ability of a neuron to fire at intervals corresponding to the fundamental and formant frequencies of a stimulus (Hall, 2007). In other words, phase-locking neurons fire at finite intervals of time corresponding to the period of the signal. The period of the stimulus is equivalent to the reciprocal of the fundamental frequency. For example, if the fundamental frequency of a stimulus is 250 Hz, the stimulus has a period of 4 ms ($1000 \text{ ms}/250 \text{ Hz}$). If a neuron has the ability to phase-lock, it would fire every 4 ms in response to a 250 Hz tone, as seen in Figure 7 below.

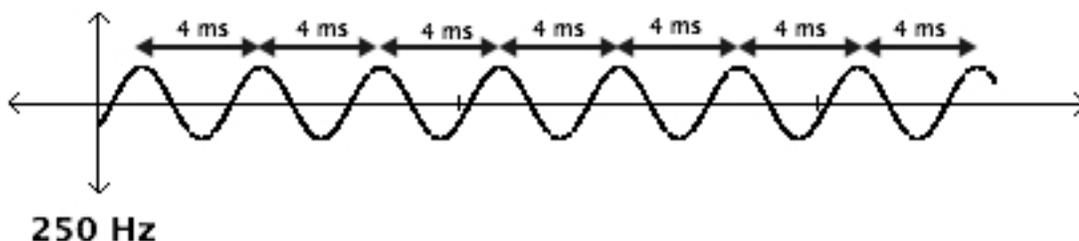


Figure 7. Phase-locking response pattern to a 250 Hz stimulus.

As previously mentioned, neural phase-locking occurs at all levels along the auditory pathway. However, as you ascend along the auditory pathway, there is a decline in both the proportion of neurons that have these phase-locking abilities, as well as the highest frequency at which phase-locking can be detected (Burkard et al., 2007). At the level of the auditory nerve, the synapse mechanisms and the filtering abilities of the hair cells become the source of this problem. This leads to deterioration in the consistency of timing of the action potential in response to the waveform. At the level of the auditory nerve, phase-locking can occur in response to stimuli up to 5000 Hz. However, at the level of the brainstem where the FFR is generated, phase-locking can only occur in response to frequencies up to 1500 Hz (Hall, 2007). The neural firing pattern underlying the FFR is determined by neural phase-locking. At the level of the brainstem, both envelope and TFS information below 1500 Hz are encoded primarily via neural phase locking.

Now that we have a basic understanding of speech acoustics, neural encoding and neural firing patterns, we can proceed to discussing the FFR and its application in understanding subcortical speech encoding in greater detail.

What is the FFR?

The FFR is a scalp-recorded AEP that reflects sustained, phase-locked neural activity of auditory neurons in the rostral brainstem in response to an auditory stimulus. (Burkard et al., 2007; Kraus & Nicol, 2005). Although a fairly recently established evoked potential, mentions of the phase-locking patterns date back to as early as the 1930s.

History of the FFR

In 1930, researchers Wever and Bray investigated the auditory system of the cat to examine the role of the cochlea and auditory nerve in neural encoding of auditory stimuli. Cats were chosen as subjects because previous studies had demonstrated that the auditory system of the cat was comparable to the human auditory system. Responses were measured in response to a range of tonal stimuli (105-5200 Hz) using several electrodes placed on the exposed feline auditory nerve. The resulting response contained a waveform that was directly related to the frequency and intensity of the stimulus (Wever & Bray, 1930). Further, Wever and Bray discovered that responses were localized in the nervous tracts, indicating that the response was not originating in the cochlea.

For several decades, this finding was not further examined. However, in 1968, James T. Marsh and Frederick G. Worden re-examined this discovery. Marsh and Worden replicated the experiment carried out by Wever and Bray (1930), by recording auditory nerve activity in response to pulsed tones using an array of electrodes implanted along the auditory pathway. The findings from this study were similar to those obtained by Wever & Bray (1930). Specifically, the morphology of the evoked response was a sine wave, similar to the test stimulus used, and lasted for the duration of the stimulus (Worden & Marsh, 1968). As this recorded activity mimicked the frequency patterns of the stimulus, Worden and Marsh (1968) chose to label it the FFR. They also went on to clearly establish a difference in the response between the FFR recorded at the cochlear nucleus (CN) and the cochlear microphonic (CM), which is an excitatory reaction of the hair cells in the cochlea recorded at the round window (RW). This difference can be

clearly seen in Figure 8 below, which depicts the monaural tracings from both the RW and CN. The responses seen at the level of the round window look to be an exact replica of the incoming signal: a sine wave with a period directly reflecting the frequency of the signal. This replica is an automatic reaction of the system and not representative of neural firing. However, at the level of the CN, although the response shows neural firing according to the period of the signal, it is not an exact replica of the incoming signal, and instead is a more jagged response of the auditory brainstem.

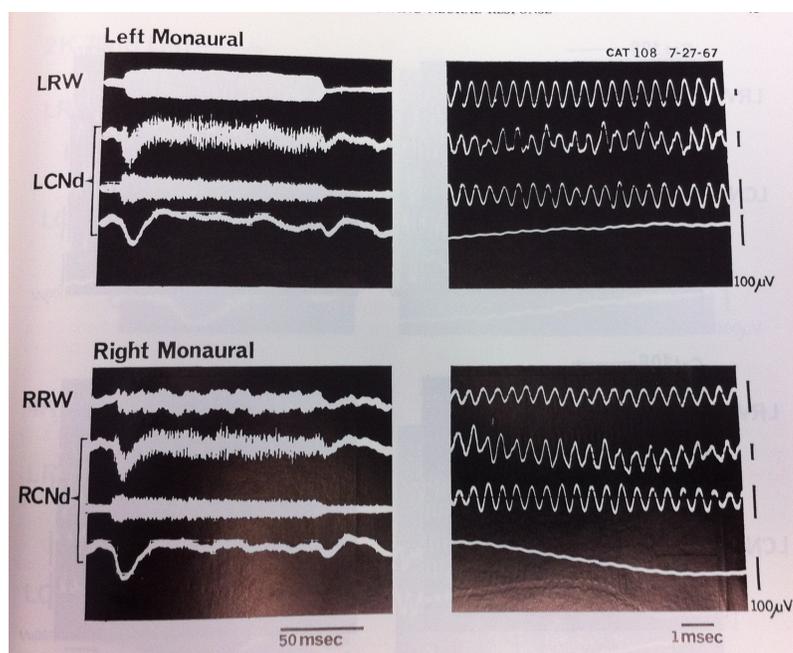


Figure 8. Monaural tracings from the round window and cochlear nucleus for both the left and right side. The stimulus used to elicit these responses was a 2000 Hz tone presented at 60 dB SPL. Recordings from the RW represent the cochlear microphonic, whereas recordings from the CN represent the FFR. From top to bottom, the three CN tracings are responses to: 1) a broad-band recording [1 Hz to 10000 Hz]; 2) a band-pass recording [1900 Hz to 2100 Hz]; and 3) a band-pass recording [1-40 Hz] (Marsh & Worden, 1968).

From their tracings the researchers noted that the two factors that affected the recording of the FFR were stimulus frequency and intensity. They found that when examining the broad-band recordings, the resemblance of the FFR to the 2000 Hz stimulus tone was obvious, whereas it was not as clear in response to the 4000 Hz stimuli. They also

discovered that the higher the intensity of the stimulus, the greater the amplitude of the response. Based on these findings, Marsh & Worden (1968) concluded that the maximum frequency range for the FFR at a stimulus intensity of 80 dB SPL is approximately 500 Hz to 5000 Hz at the level of the CN. This can be seen in Figure 9, which demonstrates the changes in the FFR as a function of the frequency and intensity of the stimulus.

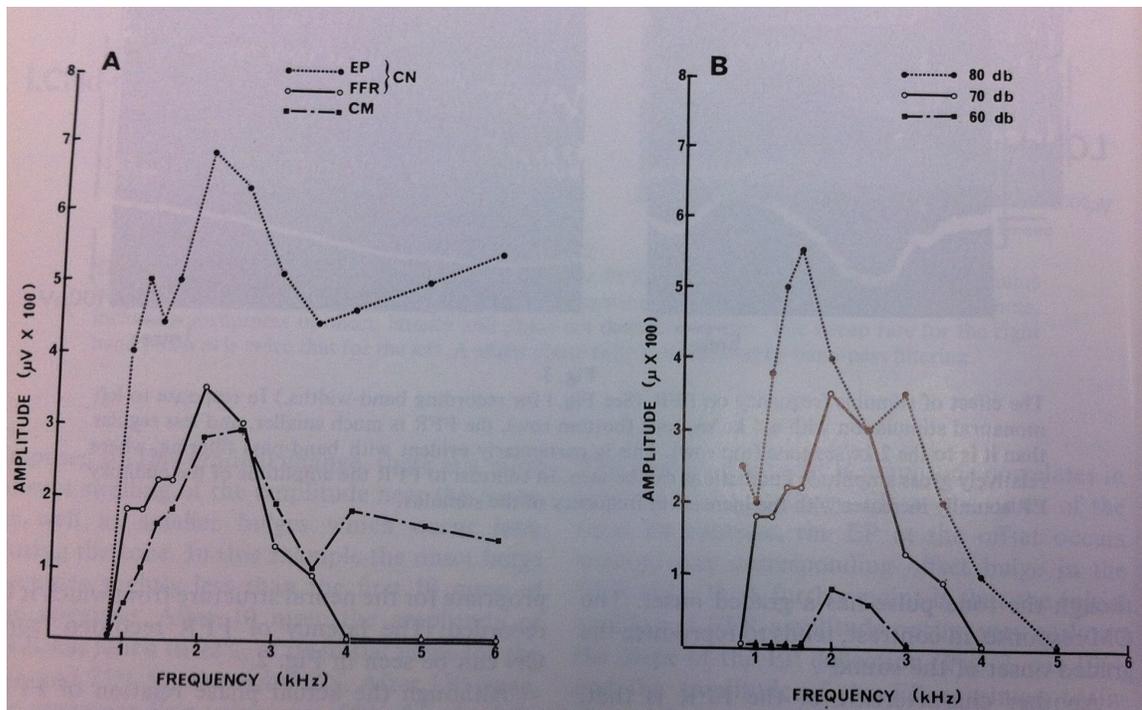


Figure 9. Amplitudes of the FFR as a function of frequency and intensity of the stimulus (Marsh & Worden, 1968).

Specifically, Panel A in the above figure depicts the amplitude of the FFR (solid black line) as a function of stimulus frequency. The highest amplitude of the FFR occurs between 500 and 4000 Hz; the amplitude drops dramatically at frequencies thereafter. Panel B represents the differences in the amplitude of the FFR as a function of stimulus intensity. The response with the greatest amplitude was elicited by the highest intensity stimulus (80 dB SPL).

Another question Marsh and Worden (1968) investigated was whether the FFR response properties changed depending on the anatomic level in the auditory system at which the potential was recorded. While neural activity could be recorded in response to stimulus frequencies as high as 5000 Hz at the level of the CN, neural activity could only be measured at 1000 Hz and below at the level of the inferior colliculus (Marsh & Worden, 1968). Furthermore, the authors found that responses recorded at the lateral lemniscus (LL) showed greater amplitude during contralateral stimulation as compared to ipsilateral. Overall, the findings from Marsh and Worden (1968) confirmed the findings from Wever and Bray (1930) and determined that the FFR is indeed an electrophysiologic response that can be recorded up to the level of the inferior colliculus (Marsh & Worden, 1968).

In 1973, Moushegian, Rupert and Stillman followed-up Marsh and Worden's 1968 study in order to determine whether the FFR could be measured via scalp electrodes in human listeners. These researchers recorded neural activity in response to pure tone stimuli from five normal hearing subjects. Recording electrodes were placed at each earlobe, the vertex, and on the leg of each subject. The high-pass cut-off filter was set to 200 Hz and the low-pass cut-off filter was set at 3500 Hz with a drop-off of 6 dB per octave. The rate of presentation was 4/s and 999 sweeps were obtained per trial. All tones were presented binaurally. The low frequency tonal stimuli below 2000 Hz consistently evoked neural responses with peaks occurring at intervals equal to the period of the stimulus, reflecting neural phase-locking (Moushegian et al., 1973). This evoked neural activity was similar to the FFR recorded in cats by Marsh and Worden (1968). Moushegian et al. (1973) demonstrated that phase-locking was more robust in response to

lower stimulus frequencies. For example, they found that a 250 Hz signal elicited a response with peaks occurring every 4 ms, a 500 Hz signal elicited a response with peaks every 2 ms, and so on. These examples can be seen in Figure 10 below in waves E and F. Thus, results from the study confirmed that the FFR could be recorded in human listeners, and reflected the underlying neural phase-locking to lower stimulus frequencies.

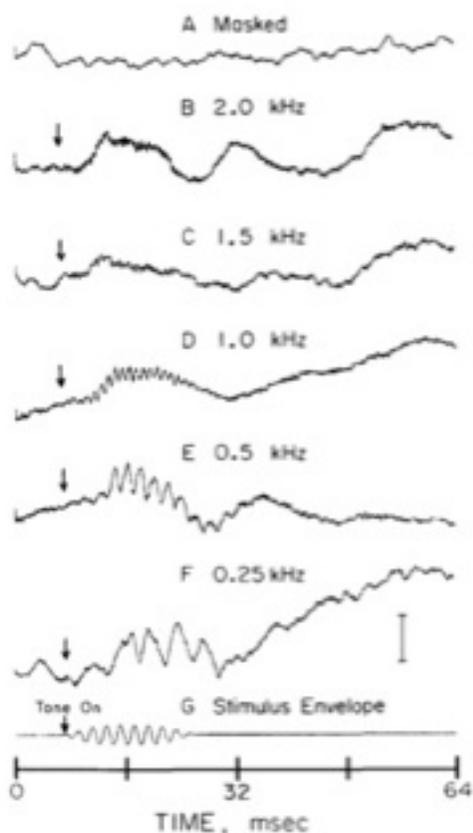


Figure 10. FFR response waveforms to different frequency tone burst stimuli (Moushegian et al., 1973).

Neural Generators of the FFR

The origin of the FFR has been a point of debate. When the FFR was first discovered by Wever and Bray in 1930, one of the goals of their initial study was to determine whether the source of the response was the cochlea or the auditory nerve. As

the responses were localized in the neural tracts, Wever and Bray (1930) concluded that the response was of neural, rather than cochlear, origin. Marsh, Worden and Smith (1970) have provided several pieces of evidence further supporting the neural origins of the FFR. Foremost, the latency of the FFR (around 5-10 ms) is appropriate to that of a neural response. Additionally, the response has a clear and abrupt onset with a discrete threshold. Furthermore, the response is so narrowly localized, it cannot be measured even millimeters away. Lastly, the response disappears under anoxia, whereas the cochlear microphonic—a cochlear response—remains (Marsh et al., 1970).

Several additional experiments have confirmed the neural origins of the FFR. Marsh, Worden and Smith (1970) conducted a two-part experiment on cats to determine whether the FFR was truly neural in origin. In the first experiment, electrodes were placed in two locations: the round window to record the cochlear microphonic; and the cochlear nucleus to record the FFR. The auditory nerve was severed and an auditory stimulus was presented monaurally. After the 8th nerve was severed, the FFR was abolished even as the CM remained. This outcome provided strong evidence in favor of a neural origin for the FFR. This finding was followed-up by a second experiment. In the second study, the left cochlear nucleus was cooled 6 degrees by the insertion of a cryoprobe, which led to reversible blocking of the FFR response. The FFR at the left cochlear nucleus was temporarily abolished during the disruption caused by the cooling, but then recovered when the ear returned to resting body temperature, indicating that the FFR is influenced by temperature change. On the other hand, the cochlear microphonic was not affected by the temperature change. Collectively, these results demonstrated that

the FFR, which is sensitive to changes in temperature, is of neural origin (Marsh et al., 1970).

A few years later, in 1974, researchers Marsh, Brown and Smith examined the distribution, symmetry, and frequency range of the FFR evoked at different brainstem auditory nuclei in two experiments. The brainstem auditory nuclei examined were: 1) central nucleus of the IC, as well as the fibers and dorsal and ventral nuclei of the LL; 2) central portion of the trapezoid body; and 3) lateral and medial superior olivary nuclei (LSO & MSO). For the first experiment, the FFR was recorded using an array of 50 penetrating electrodes placed at the different brainstem auditory nuclei (IC, LL, LSO & MSO) in 22 cats. Stimuli used consisted of pure tone bursts presented at 70-80 dB SPL. The FFR was successfully recorded from all sites of the brainstem used in the study. This finding indicates several things: 1) The FFR can reliably be recorded from all brainstem nuclei in the regions explored; 2) The FFR is large enough to be seen clearly in response to a single stimulus and does not vary from stimulus to stimulus; 3) The FFR from most nuclei is symmetrical (i.e., stimulation of the left and right ear produces an FFR of relatively equal amplitude), and 4) The binaural response is larger in amplitude than the sum of the monaural responses, indicating binaural summation. Because the FFR could be recorded even at the level of the MSO and LSO, it was deduced that the conduction of the FFR to the ipsilateral LL and IC must be via the synaptic relay of the superior olivary complex (SOC). This finding also led researchers to believe that the projection of the FFR to the contralateral LL and IC is by the direct fibers that bypass both the ipsilateral and contralateral SOC, leading the researchers to further investigate this assumption.

In the second experiment, only two cats were used and electrodes were placed at the same 50 locations as in the first experiment. Additionally, a pair of lesioning electrodes were placed along the sagittal plane in the posterior and anterior portions of the MSO. The FFR was recorded and analysis was conducted to determine the effect of the two added electrodes in comparison to the recordings from the first experiment. They found that the lesioning of the SOC affected the FFR in the IC and the dorsal nucleus of the LL only on the side of the lesion, and only when the stimulus was presented to the ipsilateral side.

Results of these experiments provided evidence of the major pathway for the conduction of the FFR, which travels from the cochlear nucleus to the LL and IC. This primary pathway to the contralateral LL and IC is direct by way of the fibers from the cochlear nucleus, which terminates in the ventral nucleus of the lateral lemniscus (VNLL), dorsal nucleus of the lateral lemniscus (DNLL), and IC. This was concluded based on the finding that the lesion of the SOC had no effect on the conduction of the FFR to the contralateral LL and IC, regardless of whether the lesion was on the ipsilateral or contralateral side of the stimulus presentation.

In 1975, Smith, Marsh and Brown compared latencies measured in direct recordings obtained at the level of the brainstem auditory nuclei in cats to the latencies measured in the scalp-recorded FFR recorded in human listeners. FFRs were recorded in response to tone bursts ranging from 60-70 dB SPL presented binaurally. Recordings were obtained from scalp electrodes in humans, and depth electrodes placed in several brainstem nuclei including CN, SOC and IC in cats. Figure 11 (below) shows the comparison of the FFR recorded from the vertex of the human and the vertex of the cat.

Panel A shows the FFR recorded at the human vertex and panel B shows the FFR recorded at the cat vertex.

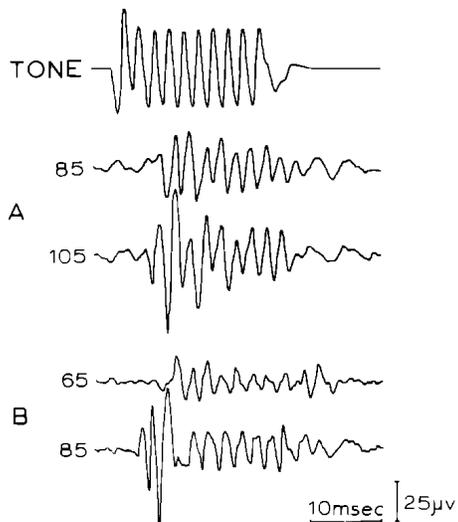


Fig 2 Effect of stimulus intensity on scalp recorded FFR responses from (A) human vertex and (B) cat vertex. Tone 500 c/sec tone burst stimulus recorded by microphone. Numbers at the left of each trace are stimulus intensity in dB (SPL). All traces are averages of 2000 responses.

Figure 11. Stimulus intensity as it affects the FFR in cat from Smith et al. (1975)

In addition, table 1 shows a comparison of the onset latencies between 18th depth electrodes and 6 scalp electrodes in 11 cats to the scalp electrodes in 9 humans.

Table 1. *A comparison of FFR latencies per electrode site in cats and humans (from Smith et al., 1975).*

TABLE I

FFR onset latencies cat and human

Recording site	Number of subjects	Mean latency (msec)*	Latency Std Dev	Latency range
Cat				
CN	4	2.0	0.29	1.6–2.4
MSO	7	3.1	0.52	2.1–4.1
IC	7	5.4	0.34	3.9–5.6
Scalp	6	5.8	0.81	5.0–6.5
Human				
Scalp	9	6.5	0.25	6.1–6.8

As evident in the table above, longer onset latencies of the FFR were seen for ascending brainstem nuclei, with the 5.4 ms average for the IC most closely approximating the 5.8 ms scalp recording from the vertex of the cat. The closeness in latency of these two points and the lack of significance in the difference between the cat and human vertex recordings indicate that the neural source of the FFR in both human and cat subjects is the IC (Smith et al. 1975).

Smith et al. (1975) followed up this finding by conducting a second experiment. In this next study, four cats were used to examine the effect of bilateral cooling of the IC on the recording of the FFR. Electrodes were placed at the left medial superior olive (LMSO), left inferior colliculus (LIC), right inferior colliculus (RIC) and at the scalp vertex. As you can see in figure 12 below, during bilateral cooling of the IC, there is a drastic decrease in the amplitude of the FFR at the level of the scalp.

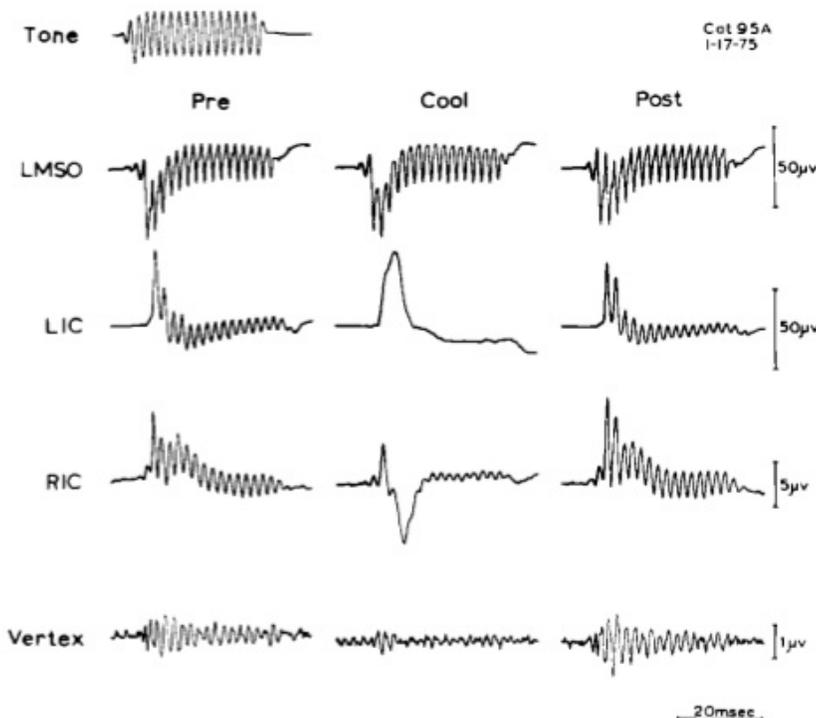


Figure 12. Recordings before, during and after cooling of the ICs at different electrode sites (Smith et al., 1975)

Recent Research Utilizing the FFR

Over the years, the FFR has been recorded to a variety of stimuli, including simple, complex, speech, non-speech, steady state, and time varying stimuli, to answer various research questions. Specifically, stimuli have included pure tones (Mousheigan et al., 1973; Smith et al., 1975; Wever & Bray, 1930), two-tone complexes (Bidelman & Krishnan, 2010), consonant-vowel constructions (Banai, Hornickel, Skoe, Nicol, Zecker... & Kraus, 2009; Cunningham, Nicol, Zecker, Bradlow, & Kraus, N., 2001; Kraus & Nicol, 2005; Plyler & Ananthanarayan, 2001), Mandarin syllables (Krishnan, Xu, Gandour, & Cariani, 2004), and words (Johnson, Nicol, & Kraus, 2005; Lorenzi, Gilbert, Carn, Garnier & Moore, 2006; Russo, Nicol, Musacchia, & Kraus, 2004). Collectively, the results of these studies indicate that the FFR is a versatile response that

can be elicited to several different types of complex stimuli. Given its ability to reflect neural phase-locking, the FFR provides a physiologic window to examine neural encoding of complex sounds such as speech at the level of the rostral brainstem. Further, single-unit studies have confirmed that phase-locking underlies the encoding of envelope and TFS cues. Hence, the FFR, which is generated by neural phase-locking, has the capacity to encode both envelope and TFS information-at the level of the rostral brainstem (Aiken & Picton, 2008; Krishnan, 2002; Skoe & Kraus, 2011; Wile & Balaban, 2007).

Recording and Measuring the FFR to Envelope and TFS Cues

The FFR is recorded to both condensation and rarefaction stimulus polarities as a first step in order to extract envelope and fine structural components of the response. A condensation polarity or “positive polarity” is associated with a positive pressure wave, and a rarefaction or “negative” polarity is associated with a negative pressure wave (Hall, 2007). A more detailed discussion of condensation vs rarefaction pressure waves will be discussed in the technical parameters section. The envelope FFR is obtained by adding the condensation and rarefaction FFRs together; in contrast, subtracting the rarefaction waveform from the condensation waveform yields the TFS components of the FFR (Krishnan, 2002). The alternating polarity method relates back to the compound histogram technique first mentioned in a study by Gobllick and Pfeiffer (1969). As described by researchers Aiken and Picton (2008), the rationale of the alternating polarity stimulus stems from the effects of “half wave rectification” involved in inner ear transduction. If the stimulus polarity is inverted, the neural discharges to the rarefaction

phase occur during the condensing phase of the initial stimulus. Therefore, subtracting the period of the histogram of the inverted stimulus cancels the distortion caused by the rectification of the rarefaction as well as the discharge pattern. Because of this subtraction method, we are able to tease out the spectral, or TFS component of the FFR. In contrast, creating a sum of the FFR to both stimulus polarities yields the overall envelope of the response (Aiken & Picton, 2008).

In order to better understand the FFR-related concepts summarized above, I will describe the steps involved in recording and analyzing the FFR in response to a speech stimulus that was carried out in our laboratory. The stimulus used was the vowel /u/ which has a F_0 of 120 Hz and a F_1 of 360 Hz, lasting approximately 250 ms (see Figure 13).

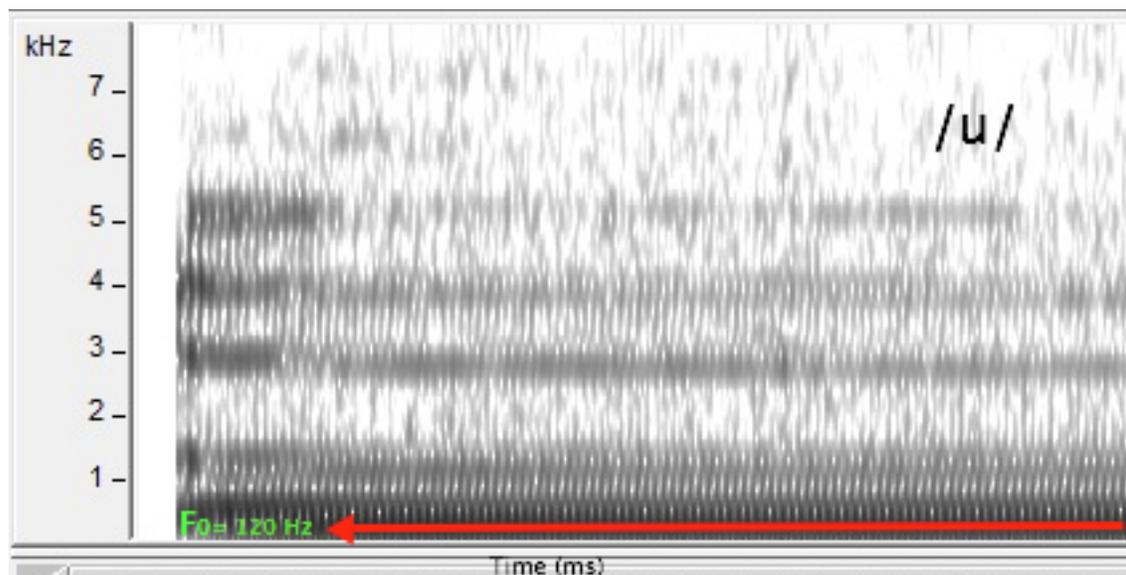


Figure 13. Spectrogram of the vowel /u/.

The FFR was recorded in both condensation and rarefaction polarities. The temporal responses that were recorded can be seen in Figures 14 & 15. As evidenced in Figure 14, repetitive peaks in the response occur in approximately 8 ms intervals from 0—240 ms, and disappear after the duration of the stimulus is complete. In this response,

the auditory neurons are firing at a set interval representing the period of the stimulus (8 ms). The recorded FFR yielded clear evidence of the neurons' phase-locking behavior, as there is a repeatable pattern of action potential firing every 8 ms.

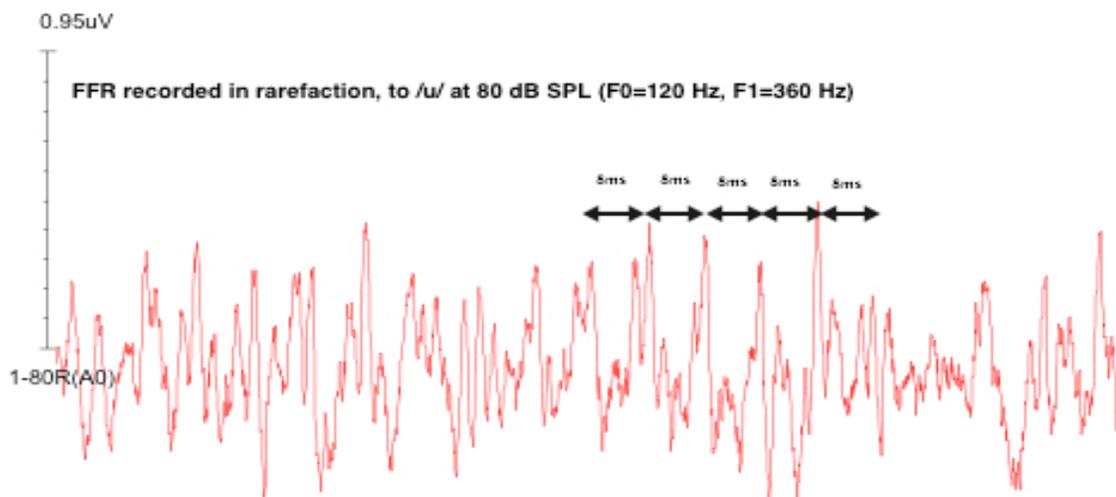


Figure 14. Rarefaction trace

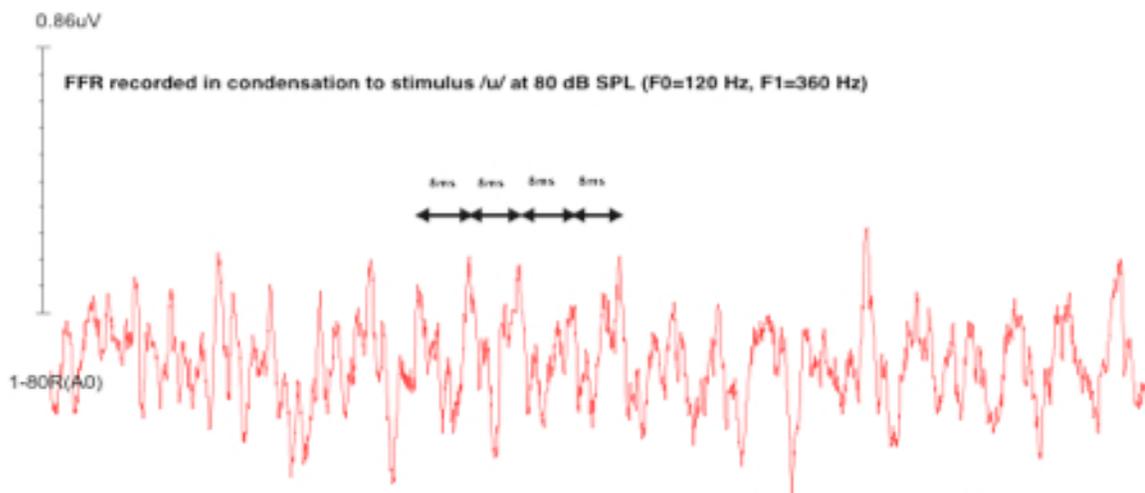


Figure 15. Condensation trace

Typically, the evoked potential system creates a separate averaged waveform to the rarefaction stimulus and a separate averaged waveform to the condensation stimulus. As described earlier, the use of these two stimulus polarities is crucial to being able to

analyze both the envelope as well as the fine structure present in the stimulus response. The addition of the rarefaction and condensation temporal waveforms provides information specific to the envelope of the FFR. The envelope of the response contains the lower frequency information, including the fundamental frequency of the stimulus (Krishnan, 2002; Skoe & Kraus, 2011). To continue on the example from earlier (Figures 14 & 15), after the FFR was recorded to the /u/ stimulus, and both rarefaction and condensation responses were averaged, the averaged waveforms were added together to obtain the envelope response (Figure 16). This was followed by a Fast Fourier Transform (FFT) in order to convert the temporal waveform into a spectral representation. The FFT, like the spectrogram, provides a graphical representation of the frequency content contained in the response. However, in contrast to the spectrogram, which plots frequency as a function of time, the FFT plots the amplitude of the response as it relates to a particular frequency. As seen in Figure 17, the FFT analysis of the summed envelope FFR reveals a robust peak at 120 Hz, the fundamental frequency of the original speech signal, confirming subcortical neural envelope encoding.

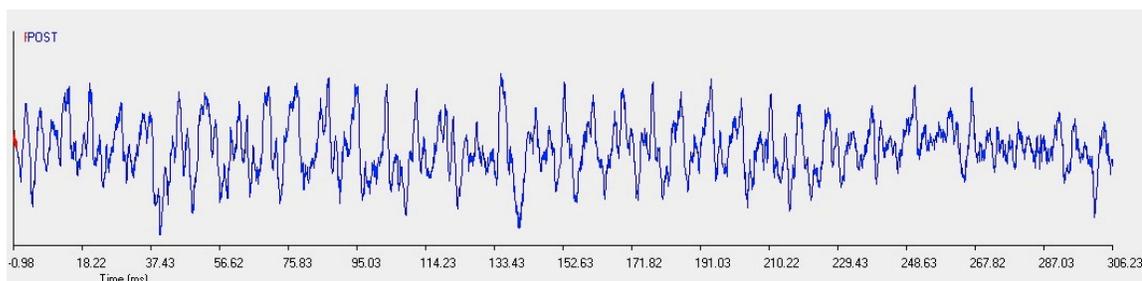


Figure 16. Added waveform rarefaction + condensation

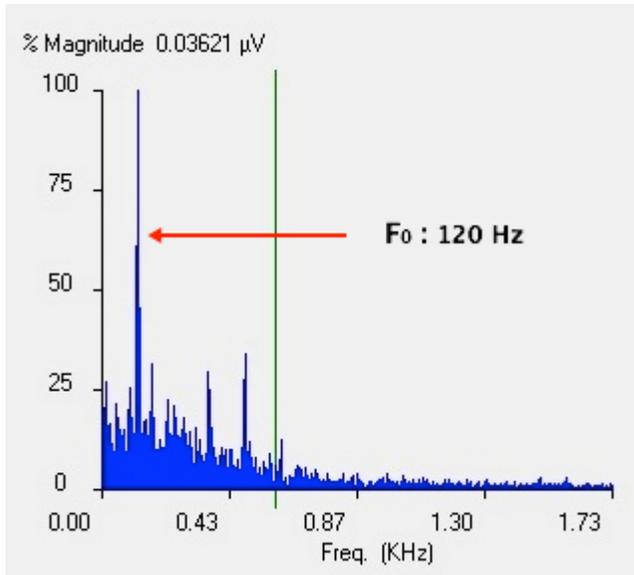


Figure 17. FFT of summed waveform

The next step in analyzing the FFR waveform is to evaluate the fine structural information of the signal. In order to do this, the averaged rarefaction wave was subtracted from the averaged condensation wave (see Figure 18). This resulting waveform gives us information about the higher frequencies present in the response, including formant frequency information (Skoe & Kraus, 2011). As before, in order to pull out the higher frequencies and fundamental frequency information from the subtracted waveform, a FFT analysis must be completed (see Figure 19). The FFT analysis yields a series of spectral peaks at the harmonic and formant frequencies present in the response, confirming that the brainstem neurons can lock on to the TFS information present in the stimulus.

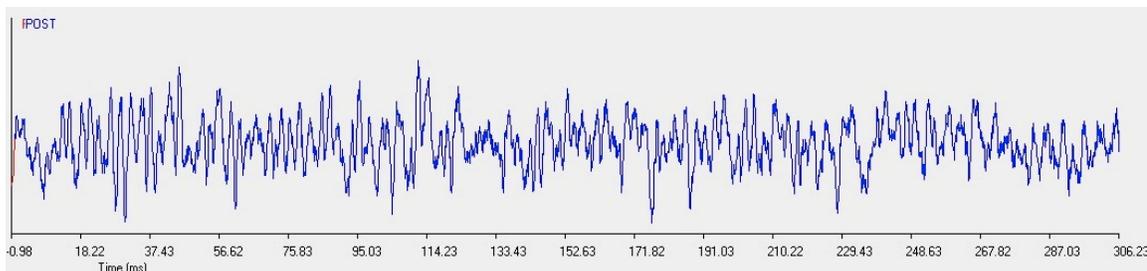


Figure 18. Subtraction waveform condensation - rarefaction

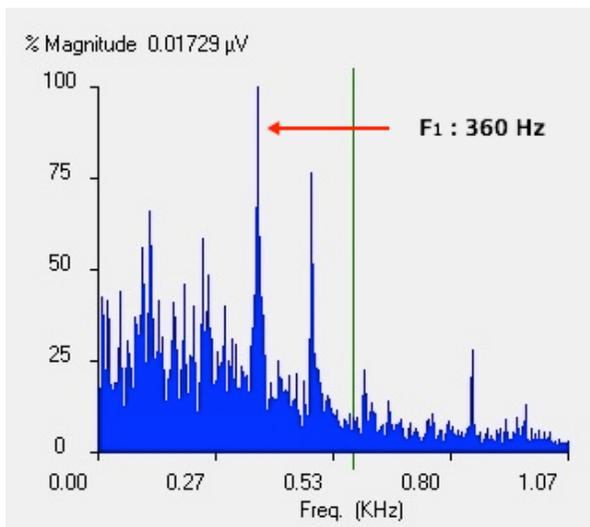


Figure 19. FFT of subtracted waveform

Just like all electroacoustic responses, the FFR can be recorded to several types of stimulus types and intensities, however there are many different technical parameters that need to be manipulated in order to maximize the response obtained by the FFR. These parameters will be discussed in the next section.

Technical Parameters

For every auditory evoked potential there are several parameters that need to be accounted for in order to maximize the amount of information that is elicited from the response. These parameters can be organized into three main categories: stimulus parameters, recording parameters, and subject parameters.

Stimulus Parameters

Stimulus parameters are settings applied to the stimulus itself. These parameters include stimulus type, stimulus intensity, stimulus polarity, and stimulus rate.

Stimulus type.

The types of stimuli used to evoke the FFR have expanded over time. As mentioned previously, the FFR was initially discovered when a pure tone stimulus was used to evoke a neural time-locked response. Since this discovery, several researchers have used different types of stimuli to study the FFR and determine what valuable clinical information it provides. These different types of stimuli can be categorized into the various categories seen in Figure 20 below.

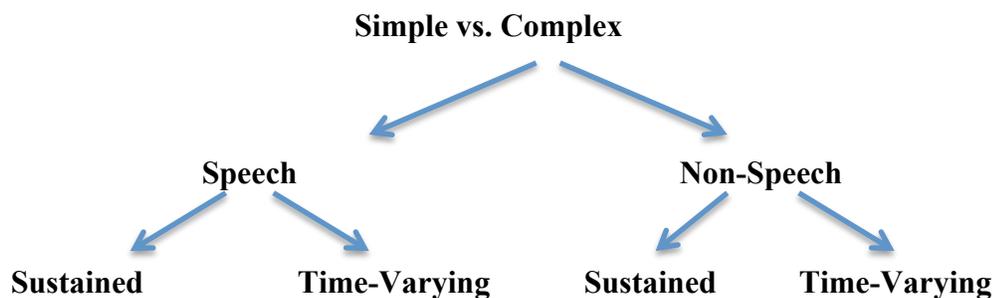


Figure 20. Diagram of simple vs. complex stimuli.

A simple stimulus is a stimulus comprised of a single frequency, such as a pure-tone. In contrast, complex signals are considered complex because there is more than one frequency being presented to the listener at a given time. Complex stimuli can be divided into two categories: speech signals (such as vowels or consonants) and non-speech signals (such as music or tone complexes). Both speech and non-speech complex stimuli can be sustained stimuli or time-varying stimuli. A sustained signal is one in which all frequency components within the signal do not change over a period of time.

An example of a complex sustained signal is a music chord. While there may be several different frequencies presented at a given time in a musical chord, the frequency components (both the F_0 and formant frequencies) involved remain unchanged over the course of stimulus duration. In contrast, a time-varying signal contains frequency components (whether it be the F_0 or formant frequencies) that change in frequency over time. The difference between time-varying and sustained stimuli can be seen in Figures 21 and 22 below. Notice that for the time-varying stimulus (Figure 21), the signal starts just below 250 Hz and then changes as a function of time. However in the sustained stimulus (Figure 22), the signal starts at 250 Hz and remains at this frequency for the duration of the stimulus.

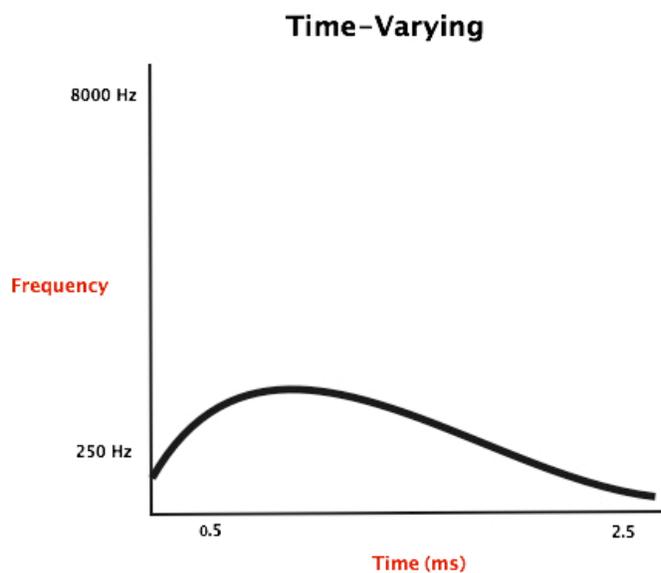


Figure 21. Time-varying stimulus.

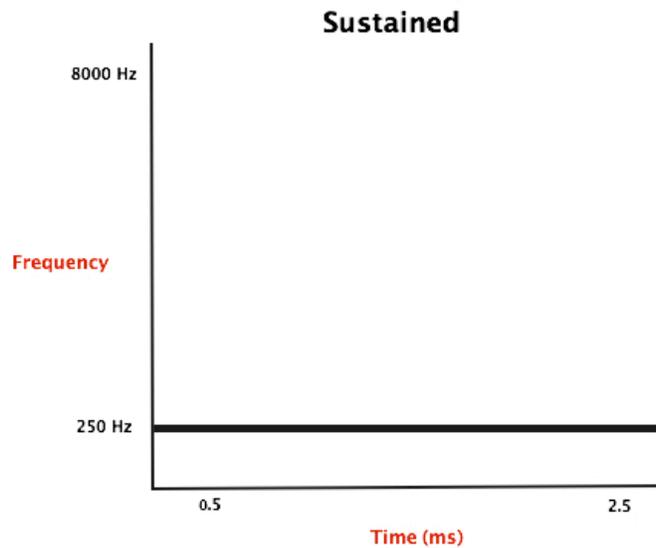


Figure 22. Sustained stimulus.

Over time, the FFR has been recorded to stimuli from each category of type of stimuli seen in Figure 20. Pure tone signals were used to elicit the FFR in several early studies discussed in the history section of this paper (Wever & Bray, 1930; Marsh & Worden, 1968; Moushegian et al., 1973). Bidelman and Krishnan (2010) used a sustained non-speech stimulus, when they studied the FFR in response to tone complexes. Krishnan and colleagues (2004) used a time-varying speech stimulus when they recorded the FFR in response to a set of monosyllabic Chinese syllables. Additionally, many studies have been conducted using sustained speech stimuli to elicit the FFR (Cunningham et al., 2001; Johnson et al., 2005; Russo et al., 2004). Sustained stimuli have a higher prevalence in the FFR literature. This could be due to the fact that this stimulus elicits a sustained response that may be easier to analyze. Regardless, the FFR can successfully be recorded to many different types of stimuli.

Stimulus intensity.

Throughout the history of this response, it has been found that stimulus intensity is an important parameter that determines whether or not a clear, replicable FFR can be recorded. Skoe and Kraus (2011) recommend a stimulus intensity of at least 60 dB SPL to record the FFR. This level is well within the conversational range of speech, as the FFR is typically recorded to investigate the neural encoding of speech signals. As early as the first study of the FFR in humans, researchers have noted that a relatively high stimulus intensity is necessary for eliciting a clear, replicable FFR. For example, when the FFR was first recorded in humans to pure tone stimuli, researchers discovered that the stimulus intensity must be at least 46 dB in order for the FFR to be recorded (Moushegian et al., 1973). As can be seen in Figure 23 below, the FFR was recorded between 5-60 dB to determine the lowest level to which the FFR could be reliably obtained. As the stimulus intensity decreases, the morphology of the FFR becomes unclear and indistinguishable. Note that the stimulus intensity that elicits the most robust response is 56 dB, however, it can still be recognized at a stimulus intensity of 46 dB.

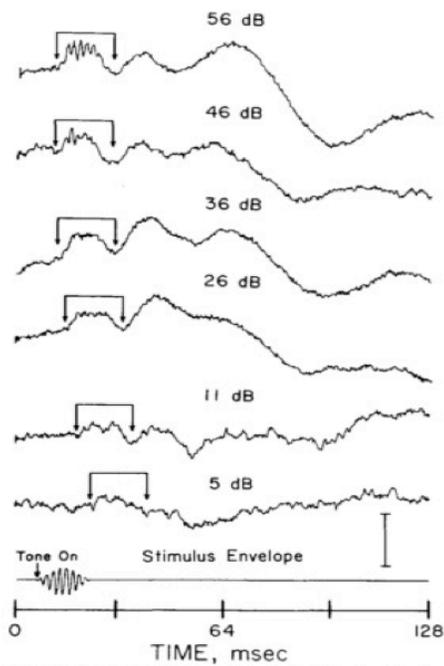


Figure 23. FFRs recorded to different stimulus intensities (Moushegian et al., 1973)

In 2002, Krishnan measured the FFR to steady-state vowel stimuli in eight normal hearing individuals in order to learn more about the brainstem's encoding of speech sounds. As a secondary research question, Krishnan also evaluated the effect of stimulus intensity on the FFR by using steady-state vowel stimuli ranging from 55-85 dB SPL. Krishnan ultimately found that although the morphology of the waveform improved with a higher intensity signal, the FFR could be recorded down to 55 dB SPL. This finding is represented in Figure 24 below. As pictured in the figure, the first two formants are clearly visible at the highest stimulus intensity. As stimulus intensity decreases, so then does the amplitude of the spikes of energy at the formant frequencies. At the lowest intensity presented (55 dB SPL), although the formants are visible, they are only slightly larger than the surrounding noise within the response.

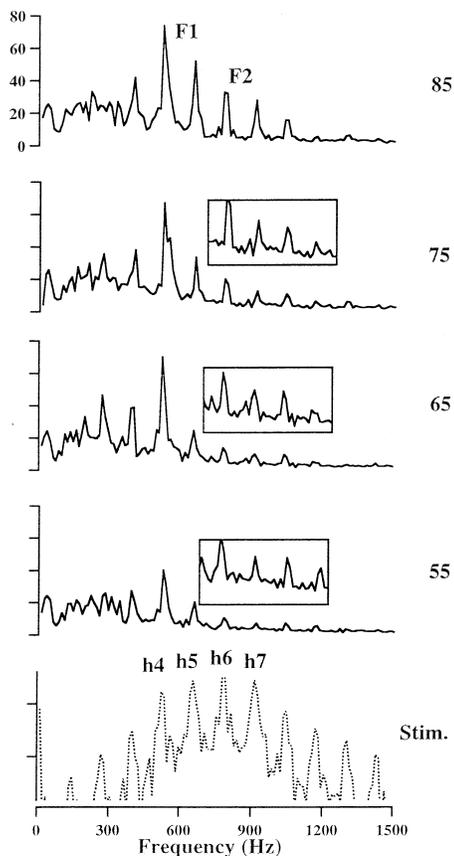


Figure 24. FFRs recorded to several different stimulus intensities (Krishnan, 2002)

A few years later, researchers Akhoun, Gallego, Moulin, Menard, Veuillet, Berger-Vachon, & Thai-Van (2008) conducted a study examining the effects of stimulus intensity on the FFR. In their study, researchers recorded the FFR of 11 normal hearing subjects using a consonant-vowel stimulus, /ba/. The subjects had hearing thresholds of 20 dB HL or below at all test frequencies. The stimulus intensities tested ranged from 0 to 60 dB SL and were increased in 10 dB steps (see Figure 25). They found that in the grand mean waveforms, well-defined FFRs were elicited from all subjects at 60 dB SL. Similar to the findings of the previous studies, Akhoun et al. (2008) reported that the FFR was not present for stimulus intensities below 50 dB SL for all subjects, and sometimes not reproducible among subjects at 60 dB SL.

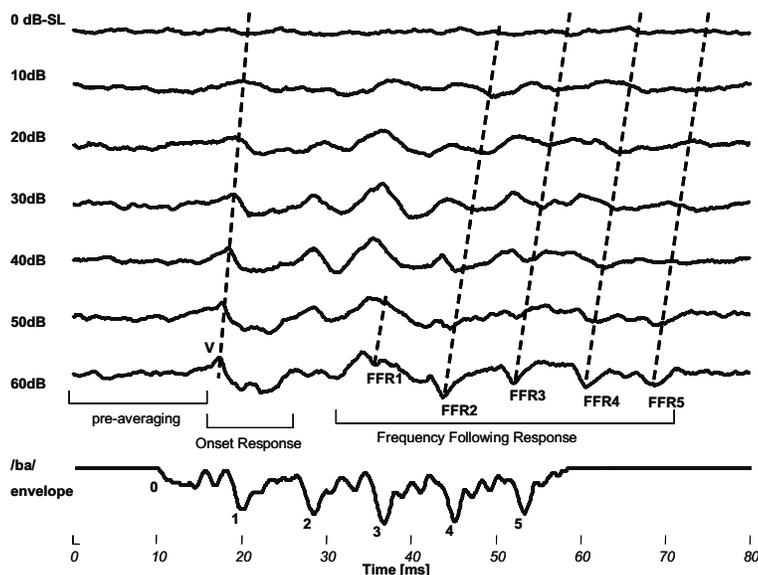


Figure 25. FFR recorded to several different stimulus intensities (Akhoun et al., 2008).

Collectively, the results of these studies have indicated that a stimulus intensity of 60-80 dB is necessary to elicit a clear and reliable FFR waveform. In the present study, we will be using a stimulus intensity within this recommended range to elicit the FFR.

Stimulus polarity.

The FFR is typically recorded to alternating, rarefaction, and condensation stimuli. Rarefaction is a stimulus polarity that initially causes the pressure wave-front of a transducer to move away from the eardrum, while condensation is a stimulus polarity that initially causes the pressure wave-front of a transducer to move toward the eardrum (Hall, 2007). Movement of a pressure wave away from the eardrum, or in a negative direction, is called a rarefaction polarity. Movement of a pressure wave towards the eardrum, or in a positive direction, is called a condensation polarity. By definition, alternating polarity stimuli alternates between rarefaction and condensation between successive trials. By recording the FFR to both stimulus polarities, the clinician is able to obtain different

pieces of information about the frequency content of the FFR that is otherwise not evident in the initial temporal waveform. As discussed earlier, the summation of the rarefaction and condensation recordings provides the clinician with information regarding the envelope or fundamental frequency information of the response. On the other hand, subtraction of the condensation waveform from the rarefaction waveform provides fine structural information regarding the formant frequencies present in the response.

The alternating polarity method relates back to the compound histogram technique first mentioned in a study by Goblick and Pfeiffer (1969). As described by researchers Aiken and Picton (2008), the rationale of the alternating polarity stimulus stems from the effects of “half wave rectification” involved in inner ear transduction. If the stimulus polarity is inverted, the neural discharges to the rarefaction phase occur during the condensation phase of the initial stimulus. Therefore, subtracting the period of the histogram of the inverted stimulus cancels the distortion caused by the rectification of the rarefaction as well as the discharge pattern. Because of this method, we are able to tease out the spectral components of the response, as well as sum both waveforms to get an overall envelope of the response (Aiken & Picton, 2008).

Many researchers have studied the FFR by using this alternating polarity method (Akhoun et al., 2008; Russo et al., 2004; Wile & Balaban, 2007). By doing so, these studies were all able to investigate neural encoding of both the fundamental frequency and the formant frequency information in the response.

In his 2002 study, Krishnan recorded the FFR in response to vowel stimuli presented in alternating polarity in eight normal hearing listeners. He went on to subtract the condensation waveform from the rarefaction waveform of each participant’s FFR to

successfully obtain fine structural information from the response including harmonic and formant frequency information. Similarly, Aiken and Picton (2008) investigated the FFR to alternating stimuli in response to vowel stimuli. The goal of their study was to examine whether using an alternating polarity stimulus would allow for the study of the main harmonics involved with speech sounds. These investigators hypothesized that the averaged response to the stimulus would have energy at the prominent stimulus harmonics as well as energy corresponding to the overall envelope. They also hypothesized that the harmonic pattern of the stimulus would be displayed when subtracting the condensation waveform of the response from the rarefaction waveform of the response, and the envelope would be displayed when adding the two waveforms together. Rather than information regarding harmonic distribution, they found that this subtraction technique provides formant frequency information related to the stimulus. This finding supported previous findings in the literature regarding the ability of the subtraction technique to highlight formant frequency information related to the stimulus.

The findings of this study, as well as the use of the subtraction and addition technique of other studies, provide the rationale for using alternating polarity stimuli when recording the FFR in the current study.

Stimulus rate.

The choice of an appropriate stimulus rate when recording the FFR has been a long-standing topic of discussion between researchers. There are two ways to calculate the rate of the stimulus: using the inter-stimulus interval (ISI) method or the stimulus onset asynchrony (SOA) method. The ISI is the period of silence (in ms) between the

offset of one stimulus and the onset of the next. In contrast, the SOA is measured from the onset of one stimulus to the onset of the next stimulus (in ms) and thus it includes the total duration of the stimulus. ISI has an inverse relationship with stimulus rate, such that $ISI = 1 \text{ sec} / \text{stimulus rate}$. Therefore, if the rate of the stimulus is 10/s, $ISI = 1000 \text{ ms} / 10 = 100 \text{ ms}$.

According to Skoe and Kraus (2011), there are three factors that should be considered when choosing an appropriate ISI. First, the ISI should be long enough so that the response to the first stimulus has time to conclude before the onset of the next stimulus. After the stimulus is presented, the neurons enter a refractory, or resting, period after they have fired. If the next stimulus is presented before the conclusion of this refractory period, and thereby the neurons do not have enough time to recover before they have to fire again, the robustness of the response will be degraded. This degradation results when the neurons do not have enough time to recover before the next stimulus is presented and they must fire again (Hall, 2007).

The second factor to consider when choosing the ISI is in regards to the length of the averaging window. This also can be controlled for by both the ISI as well as the duration of the averaging window. The averaging window should be long enough so that each stimulus can return back to baseline before the next stimulus onset. An averaging window that is too short will cause the nerve to go into an adaptation mode, causing it to be less sensitive to the incoming stimulus. The brain recognizes the stimulus as a constant and therefore the response becomes progressively less robust.

The third factor to consider when choosing the ISI involves the stimulus rate and the AC line frequency (60 Hz). The stimulus rate should be chosen so that if it is divided

by the line frequency, the resulting number is not an integer. For example, Skoe and Kraus (2011) recommend using a rate of 10.3/s or 11.3/s rather than a rate of 10/s. This will avoid contamination from the AC line frequency during recordings.

Many studies have successfully recorded the FFR using stimulus rates in the range of 3.13/s to 3.39/s (Jeng, Hu, Dickman, Montgomery-Reagan, Tong, Wu & Lin, 2011; Krishnan et al., 2004). Studies that have investigated rate effects on the integrity of the FFR have found that presenting stimuli at high rates (such as 10.9/s or 15.4/s) will degrade the response, especially in the higher frequency range (Krizman, Skoe & Kraus, 2010). After taking these factors as well as results of several studies into account, the present study will use a stimulus rate of 3.1/s.

Recording Parameters

Recording parameters are settings within the recording apparatus that effectively capture the necessary components of the response. These may include the electrode montage, the analog band-pass filter settings, sampling rate, number of sweeps, length of the analysis window, and artifact.

Electrode montage.

The combination of two or more electrodes placed on the head forms an electrode array or montage (Hall, 2007). There are two electrode montages that have been used throughout the literature to record the FFR: the vertical montage and the horizontal montage. The most commonly used of these two is the vertical montage. This 2-channel montage involves five electrode placements: a grounding electrode located on the lower

forehead (Fpz), an active electrode located on the high forehead (Fz), and two inverting (reference) electrodes at the linked mastoids (M1 & M2) and one at the nape of the neck (C7). This montage is depicted in Figure 26 below.

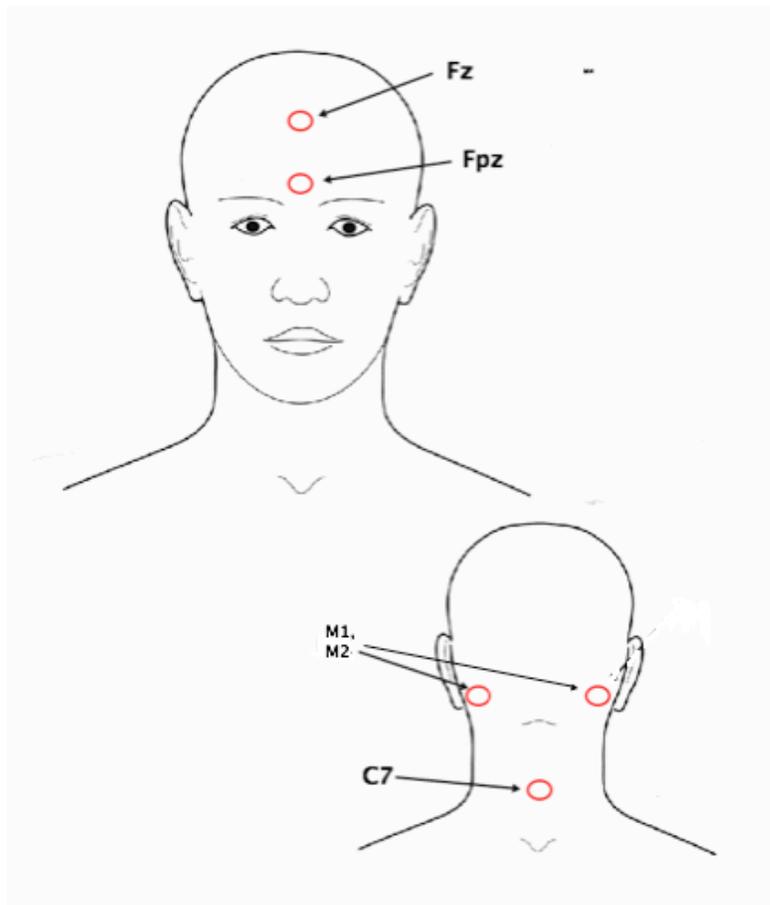


Figure 26. Vertical montage

A horizontal montage has also been used to record the FFR. While this setup is similar to that of the vertical montage, the horizontal setting requires the active electrode to be placed on one earlobe, rather than the high forehead, with the reference electrode on the opposite earlobe. This setup is depicted in Figure 27 below.

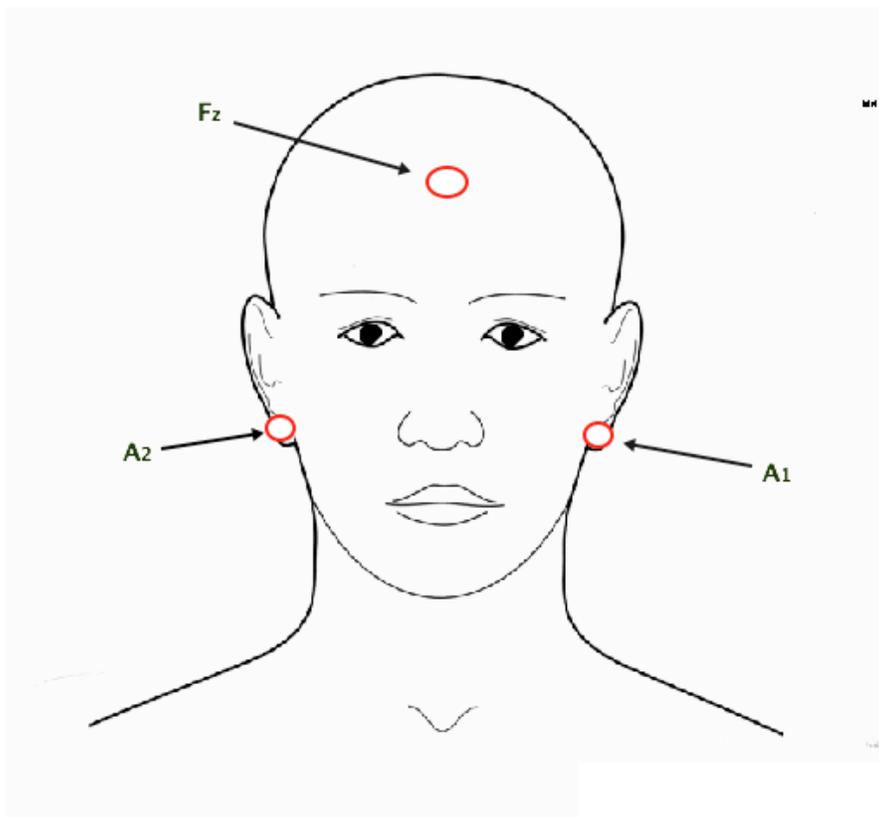


Figure 27. Horizontal montage

Galbraith (1994) used both vertical and horizontal electrode montages in his study investigating the FFR's role in encoding "missing fundamental" stimuli. Normal hearing listeners were tested in a series of two experiments: the first using a pure tone stimulus, the second using a complex missing fundamental stimulus. Two montages were employed to determine which configuration led to a more accurate and reliable FFR. Assumptions can be made based on the knowledge of the structures along the auditory pathway. Due to the location of the electrode placement for the horizontal montage, information from the peripheral level (i.e., the auditory nerve) would dominate the response. In contrast, electrodes placed in a vertical montage would capture energy of the response at higher structures along the auditory pathway. We know that the neural generators of the FFR lie in the rostral brainstem, namely the IC and LL (Galbraith,

1994; Smith et al., 1975), therefore a vertical montage should capture more accurate response information pertaining to the FFR. Galbraith reported the recordings from the horizontal montage yielded earlier response latencies in comparison to the latencies recorded from the vertical montage. For example, the average latency of the FFR was 2.15 ms when recording using a horizontal montage, and 5.23 ms using the vertical montage. In addition to latency differences of the FFR, the spectral characteristics of the FFR were different for each montage. The vertical montage yielded a clear and well-defined five-component response, whereas the horizontal montage yielded no such organization. These longer latencies and spectral characteristics, as captured by the vertical montage, relate to the central auditory structures such as the IC and LL, and thus Galbraith's findings support use of the vertical montage in recording the FFR. Therefore, it is more clinically relevant to use a vertical montage rather than a horizontal montage in the current study.

Analog band-pass filter settings.

Analog filtering is a technique used to reduce the amplitude of unwanted electrical noise without altering the energy present in the desired neural response that is being recorded (Hall, 2007). Band-pass filters used for electrophysiology recordings should be chosen so that the energy present in the response is well within the selected high-pass and low pass cutoff frequencies. There are four main types of analog filters: low-pass, high-pass, band-pass, and band-reject filters. These filters allow energy to pass at certain frequencies while rejecting energy at other frequencies (Hall, 2007). For example, low-pass filters reject higher frequency energy and allow energy below the

cutoff frequency to pass through. In contrast, high-pass filters reject energy below the cutoff frequency and allow higher frequency energy to pass through. Band-pass filters reject energy below a certain cut-off point and above a higher cutoff point, allowing only energy between these two cut-off frequencies to pass through (Hall, 2007). Band-pass filters are commonly used to record AEP measurements.

For the purposes of this study, the settings, or cut-off frequencies for these filters are determined by two factors. The first factor is the stimulus we are using. It is important to be sure that the frequency range of the stimulus is well within the cut-off frequencies of the band-pass filter. For example, if the vowel /u/ serves as the stimulus, it would be necessary to ensure that the fundamental frequency is higher than the low-pass filter in the band-pass filter setting, and that the highest formant frequency is lower than the low-pass filter setting. Thus if the fundamental frequency of the vowel /u/ is 120 Hz, the high-pass filter setting must be lower than this fundamental frequency.

The second factor concerning the present study is the phase-locking abilities at the level of the auditory system where we are eliciting a response. Because the auditory structures at the brainstem level have lower limits for phase-locking (around 1500 Hz), the analog band-pass filter must be set to encompass this lower frequency range in order to preserve all of the lower frequency phase-locking properties of the auditory nerve. In the current study, a band-pass filter of 100-3000 Hz will be used.

Sampling rate.

The sampling rate determines how many times the analog signal is digitally sampled by the recording system (Skoe & Kraus, 2011). The basis around choosing an

appropriate sampling rate is directly related to the Nyquist frequency, defined as the highest frequency present in the stimulus (Skoe & Kraus, 2011). The Nyquist theorem then states that an appropriate sampling rate is one that is at least two times higher than the highest frequency present in the stimulus (i.e., the Nyquist frequency). The rationale for this sampling rate is to preserve as much of the temporal waveform as possible, and to produce a response that represents the analog signal as close to the original as possible. Increased sampling improves the clarity of the recorded response and is therefore recommended by researchers (Skoe & Kraus, 2011). In the present study, a sampling rate of 6,000 will be employed.

Number of sweeps.

The number of sweeps, or trials, can also be defined as stimulus repetitions (Hall, 2007). When recording an AEP, a higher number of sweeps is recommended in order to increase the signal-to-noise ratio (SNR), and therefore increase the clarity and robustness of the recorded waveform. During the signal averaging process of the recording, there is a maximum response of interest (e.g., the phase-locking response of the brainstem structures responsible for the FFR) that is embedded within the surrounding EEG noise of the subject. As you increase the amount of sweeps, the random and non-uniform EEG noise is cancelled out over time, while the response of interest is summed and enhanced. This signal averaging process theoretically increases the SNR and makes the analysis of the response much easier. The range of stimulus repetitions most often used in FFR studies has been found to be between 1000 and 2000 sweeps (Aiken & Picton, 2006; Dajani et al., 2005; Krishnan, 2002; Russo et al., 2004). However, a higher number than

this range can only improve the quality of the response waveform. Therefore, it is recommended to use between 2000 and 3000 sweeps to record the FFR (Skoe & Kraus, 2011). This provides enough sweeps to enhance the FFR, without adding too much to total test time. In the current study, a total number of 2000 sweeps will be used per trial.

Length of the averaging window.

As a general rule for the recording of AEPs, the length of the averaging window should be long enough to encompass the response of interest in all test conditions (Hall, 2007). For the FFR, the length of the averaging window should be long enough to include the pre-stimulus baseline, the response period, and the post-response period (Skoe & Kraus, 2011). The pre-stimulus period allows for the clinician to clearly see where the FFR begins. A good pre-stimulus length is approximately 40 ms. The response period depends on the length of the stimulus being used. The post-stimulus period allows for the response to return back to baseline. A good post-stimulus length is approximately 10 ms. With these three components taken into account, the stimulus used in the present study is lasting approximately 260 ms. In the present study, we will use a speech stimulus /u/ which has a total duration of 260 ms, with a post stimulus period of ~40 ms.. In order to eliminate any onset responses from the data analysis, the first 20 ms of the response will be discarded during analysis; hence there is no need for a pre-stimulus period. Therefore, the total length of the analysis window will be 300 ms.

Artifact.

Artifact can be defined as a contamination of a recorded neural response (Hall, 2007). There are several different kinds of artifact for any AEP. Specifically in terms of

the FFR, four main types of artifact need to be controlled for. These include: electrical noise, myogenic artifact, the cochlear microphonic, and stimulus artifact.

Electrical noise is the result of the line noise usually emanating from an electrical socket. This is typically within the range of 50-60 Hz. This artifact can be controlled for by using the notch filter within the recording software, and/or by using an electronically-shielded room.

Myogenic artifact, or muscle artifact, is caused by movement of the participant during the recording. This artifact is particularly worrisome because the amplitude of the response can be much larger than the brainstem response. To eliminate this artifact from contaminating the response, artifact rejection criteria of ± 20 to $\pm 75 \mu\text{V}$ is recommended (Skoe & Kraus, 2011). However, smaller muscular movement may continue to go undetected. Therefore, in attempt to control this remaining myogenic artifact, the participants involved in the present study will be instructed to relax and be still with their eyes closed, to prevent myogenic artifact.

The CM is the instantaneous obligatory response of the outer hair cells immediately following the presentation of an auditory stimulus. As discussed previously, as the neural generators of the FFR originate at higher levels of the auditory pathway (within the rostral brainstem), the latencies associated with these generators are much later than the instantaneous response of the CM (the FFR has a latency of approximately 6-10 ms following the initial presentation of the stimulus). This artifact, while it cannot be avoided, is controlled for by the informed analyses of the response by an audiologist who can determine the difference between the CM and the FFR in the response waveform.

Stimulus artifact is electrical artifact that is produced by acoustic stimulus transducers (Hall, 2007). These transducers produce an electromagnetic field that generates electrical activity. Often during AEP recordings, transducers and their resulting electromagnetic fields are located near, and consequently picked up by electrodes. As a result, this electrical activity is included in the overall response. In order to control for this artifact, the use of electromagnetic shielding earphones is recommended. Furthermore, audiologists setting up patients for testing should make a conscious effort to separate transducer wires from electrode wires (Hall, 2007).

In the present study, the FFR will be recorded in an electronically-shielded room with the 60-Hz line filter on. It will also be recorded using electromagnetic shielding transducers, setting the artifact rejection criteria to be +/- 20 to +/- 75 μV , and encouraging the participant to relax or even sleep if possible.

Subject Parameters

Subject parameters are controlled for by choosing the proper subject base as well as subject protocol. The subject parameters that can affect the recording of the FFR are: attention, sleep state, and age/maturation effects.

Attention.

Subject state is a parameter that has been discussed throughout the FFR literature. The FFR has been recorded to subjects participating in attention tasks as well as to subjects presented with no attention tasks. Musacchia, Strait and Kraus (2008) conducted a study in which 26 normal hearing adults attended to an auditory stimulus in order to

perform a discrimination task. The study was conducted using the speech syllable /da/ in three listening conditions: 1) when listening to the sound alone and simultaneously watching a captioned video; 2) when listening to the sound and simultaneously watching a video of a male speaker saying “da”; and 3) when only watching the speaker say “da” without sound. The researchers found that the FFR could be recorded while the subjects attended to the stimulus as well as when they were not attending to the stimulus. In contrast, Galbraith and colleagues have conducted a series of studies on the effect of selective attention on the FFR (Galbraith & Arroyo, 1993; Galbraith & Doan, 1994; Galbraith & Kane, 1993). In their 1994 study, Galbraith and Doan recorded the FFR in 34 normal hearing adults using a 400 Hz stimulus as well as a complex, missing fundamental stimulus. Each stimulus was presented to one ear so that the subjects listened to both stimuli simultaneously. The subjects were instructed to listen to one stimulus and ignore the other. The specific stimulus they were instructed to attend to was dependent on the group into which they were randomly sorted. Results of the study indicated a significant interaction effect in certain stimulus conditions. Namely, more of an effect was seen for the pure tone stimuli rather than the complex stimuli. However, these effects were not found in the previous studies conducted by Galbraith and his colleagues. Although the results of these studies vary, it appears that the overall effect of selective attention can alter the signal processing and thereby the waveform of the FFR (Galbraith & Arroyo, 1993; Galbraith & Doan, 1994; Galbraith & Kane, 1993). Due to the possible negative effect of attention on the output of the FFR, it is best that the subject be relaxed throughout the duration of the recording, and not distracted by another input that would cause the subject to ignore the stimulus.

Sleep state.

Due to the threat of myogenic artifact as mentioned previously, Skoe and Kraus (2011) recommend to record the FFR while the subject is sleeping. Several researchers have supported this recommendation by successfully recording FFRs in sleeping subjects (Aiken & Picton, 2008; Dajani et al., 2005). Aiken and Picton (2008) recorded the FFR in 10 normal hearing adults using two vowel stimuli: /a/ and /i/. Subjects in this study were encouraged to sleep if possible; if they were unable to sleep, they were told to relax as much as possible. Dajani et al. (2005) recorded the FFR to vowel stimuli in seven normal hearing adults. Both studies were able to record reliable and robust FFRs for all sleeping subjects. This finding suggests that there is no significant change in the FFR waveform morphology during the sleeping state. In the present study, the subjects will be encouraged to sleep, if possible, but at minimum to relax with their eyes closed during acquisition of the response.

Age/maturation effects.

Several researchers have discussed the effects of age/maturation on the FFR. Some researchers argue that the auditory pathway in the brainstem in infants matures as young as 1 month of age. This conclusion was due to the successful recordings of the FFR in very young children. These recordings were similar to those recorded in adult subjects (Levi, Folsom & Dobi, 1994). Other FFR studies conducted on young children have suggested that the FFR is not fully developed until the age of 2 or 3 months (Jeng et al., 2011). Johnson, Nicol, Zecker and Kraus (2008) argue that the early auditory pathway is not fully developed until the age of 2 years. These conclusions have been drawn based

on results of their study showing ABR waveforms continue to change over time. Specifically, they found that the mature ABR waveform is not seen until the age of 2 years, indicating full maturation of the early auditory pathway at this time. In their study, they recorded the FFR on 104 normal hearing children between the ages of 3 and 12 years using a speech syllable /da/. They found that latencies for the younger age group (3-4 year old) were later than the latencies found for the older children. However, they did not find any differences in the morphology of the waveforms between the two groups. In contrast, Clinard and colleagues (2010) studied age effects on the FFR in middle-aged to elderly adults. More specifically, they tested 32 normal hearing adults between the ages of 22 and 77 years. FFR data were collected in response to pure tone stimuli. Results of this study indicated that with increasing age, FFR morphology and formant representation worsened. They related the decline in morphology in the higher frequency components of the FFR to the decreased phase-locking ability that accompanies aging (Clinard, Tremblay & Krishnan, 2010). In the present study, we will be recording the FFR in normal hearing adults between the ages of 20 and 35 years. Therefore, age and maturation effects will not affect our data.

Psychoacoustics: reverberation and background noise (BGN)

While it is not uncommon for listeners with sensorineural hearing loss to have decreased speech understanding ability in quiet listening environments, these difficulties are often exacerbated in adverse listening conditions. Two examples of such adverse conditions are background noise and reverberation. These conditions can make speech intelligibility difficult even for the normal hearing listener. Both reverberation and

background noise can alter the spectral components of a speech signal, resulting in a degraded speech signal. Such degradations in the speech signal cause deficits in speech understanding, which are particularly problematic for listeners with sensorineural hearing loss. While both reverberation and background noise affect the speech signal, the nature of degradation in the speech spectral components of the signal in both conditions is different. In the following section we are going to discuss the effects of background noise and reverberation on neural encoding of speech.

Background Noise

In its broadest description, noise is a complex aperiodic stimulus (Emanuel & Letowski, 2009). More specifically, there are two types of noise: non-stationary and stationary. Non-stationary noise may change in content and/or intensity over time. Examples of non-stationary noise would be multi-talker babble, street noise, and music. In contrast, stationary noise is a random complex stimulus that does not vary much over time. Examples of stationary noise would be white noise, pink noise, or speech-spectrum noise.

Effects of background noise on speech acoustics.

Vowels are typically identified by determining the difference between the formant peaks in the speech spectrum (peak-to-valley ratio) and the remainder of the spectrum (Assman & Summerfield, 2003). This difference is referred to as the spectral contrast. When background noise (BGN) mixes with a speech signal, the spectral contrast crucial to vowel identification is reduced by the presence of noise in between formant

frequencies (Assmann & Summerfield, 2003). According to Nabelek and Nabalek (1994), several parameters determine the effect of noise on speech intelligibility. These include the long-term spectrum of the noise, the intensity fluctuation of the noise over time, and the average intensity of the noise relative to the intensity of speech, also known as the signal-to-noise ratio (SNR). The signal-to-noise ratio can be defined as the ratio of some measured aspect of a signal (in this case, intensity) to some measured aspect of concurrent noise (intensity); usually expressed in logarithmic form (i.e. x dB SNR) (Emanuel & Letowski, 2009). A SNR can be positive, negative or 0. A positive SNR means that the signal is louder than the noise. For example, a +6 dB SNR means that the signal of interest is 6 dB louder than the background noise. A negative SNR means that the noise is louder than the signal. For example, a -6 dB SNR means that the signal of interest is 6 dB quieter than the background noise. A 0 dB SNR means that the signal and the noise are at equal intensity levels. For example, both the signal and the noise are at 60 dB. The degree of the SNR can determine how much of the speech spectra are distorted. A 0 or negative SNR has more detrimental effects on the spectra than a positive SNR in normal hearing listeners. The current study focuses on the effects of SNR on neural encoding of a speech signal and therefore this parameter will be the focus of the discussion.

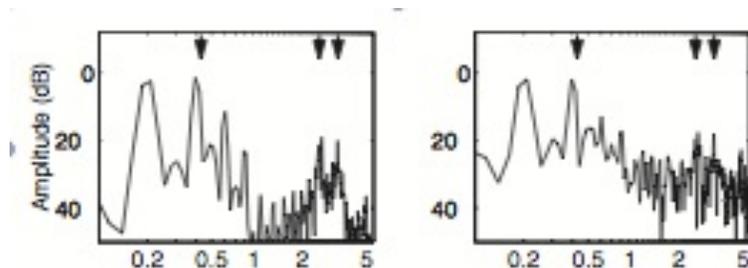


Figure 28. FFTs of vowel in quiet (left panel) vs. vowel in +6 dB SNR (right panel)

Examine figure 28 above (Assmann & Summerfield, 2003). This figure shows a comparison of two FFTs of a vowel in quiet and in the presence of background noise. The left panel demonstrates the vowel in quiet. Notice the spectral contrast of each formant peak and the energy disbursement at each frequency. The right panel demonstrates the vowel in the presence of background noise at +6 dB SNR. Notice how the spectral contrast is still maintained at the formant peaks, however the peak-to-valley ratio is decreased slightly due to the addition of noise. Although there is a change in the speech spectrum, the disruption to the spectral contrast is not enough to make the vowel unintelligible. However, as the SNR decreases, the effect of the noise on the spectral contrast worsens.

Such a loss in spectral contrast due to the addition of a background noise is illustrated in Figure 29 below, which contrasts the FFT of a vowel in quiet with that of a vowel in the presence of background noise at an SNR of 0 dB.

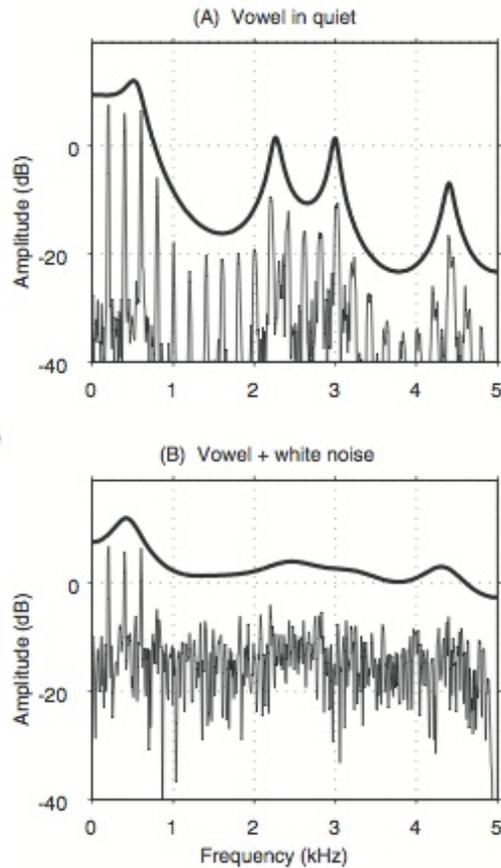


Figure 29. From Assmann & Summerfield (2003)

When a vowel is presented in quiet as seen in panel A, the spectral peaks, or formant frequencies, are well defined and clearly distinguishable. There are clear peaks visible at approximately 500, 2250, 3000 and 4500 Hz. However, with the introduction of background noise as seen in panel B, the spectral contrast is not clear at all. A slight peak can be observed around 500 Hz, however, all other formant peaks are lost within the noise. Without these spectral peaks, vowel identification becomes much more difficult.

Effects of background noise on speech understanding.

The effect of background noise on speech understanding has been evaluated using many different kinds of speech stimuli ranging from vowels (Nabelek & Dagenais, 1986;

Nabelek, 1988; Pickett, 1957) to words (George, Goverts, Festen, & Houtgast, 2010) to sentences (Duquesnoy & Plomp, 1980). However, as the present study is evaluating neural encoding of a vowel sound, this section will focus on studies examining the effects of background noise on short-segment stimuli such as vowels, consonants and diphthongs.

Vowels are typically steady-state vocalizations that are differentiated primarily by their formant frequency information. Given that background noise often alters the representation of formant frequencies, vowel sounds are highly susceptible to the influence of adverse listening conditions.

Pickett (1957) first examined the effects of background noise on speech intelligibility in 11 normal hearing listeners. Two sets of syllables were used: one set in which syllables occurred with equal probability, and the other set in which the vowels occurred approximately as often as they do in the English language. These syllables were spoken by male talkers between the ages of 23 and 33 years. Three different types of noise were used: low frequency noise, high frequency noise, and flat noise. Listeners were instructed to report which syllable they heard by either repeating the syllable or writing down the syllable. There were several different types of errors in vowel identification found in this study. These errors were made as a result of the background noise present in congruence with the speech signal. The frequency range of the noise indicated which part of the speech signal was masked out. For example, the low frequency noise masked out primarily fundamental frequency information, whereas the high frequency and flat noise masked out the higher formant frequency information. Errors also were made depending on the fundamental frequency information of the

speech signal presented. For example, a vowel such as /a/ would be more difficult to identify in the presence of low-frequency noise due to its lower fundamental frequency. A vowel with a higher fundamental frequency, such as /i/, may be easier to distinguish in the presence of low frequency noise. Pickett noted that the shifts in vowel confusions were consistent with a formant theory of perception, which states that when a noise masks one formant, the unmasked formant is the one that is correctly perceived. Figure 30 below shows FFTs of two different vowels and their respective formant frequencies. Notice how both vowels have energy at similar frequencies, however, the formants assigned to these frequencies are different.

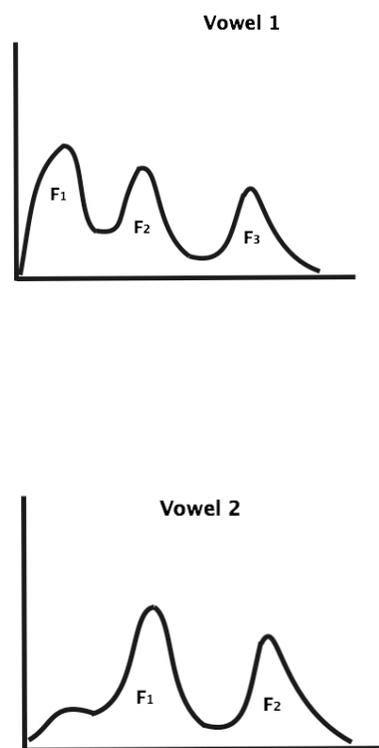


Figure 30. FFTs of two separate vowel stimuli.

Figure 31 demonstrates what happens to the frequency information represented in the FFT of Vowel 1 when low frequency background noise is presented along with the speech signal. Notice how the low frequency noise masks out the first formant frequency of Vowel 1, causing it to now look very similar to the first and second formant of Vowel 2.

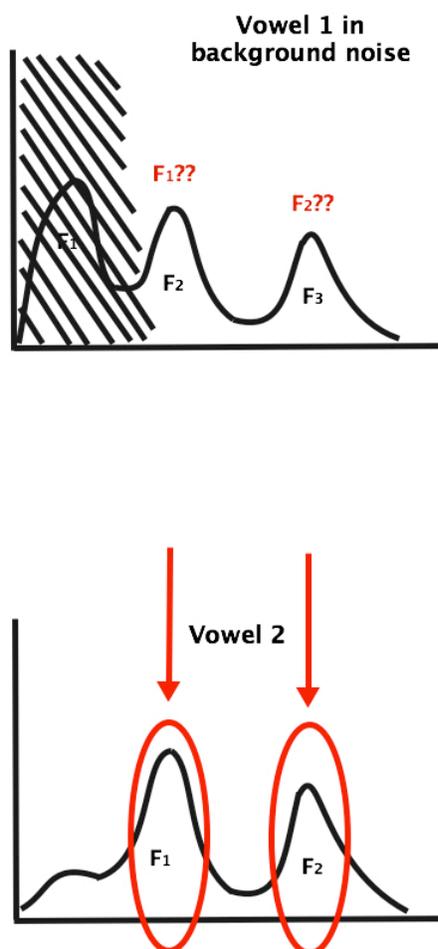


Figure 31. FFTs of a vowel in BGN (top panel) and the vowel it is perceived as (bottom panel).

This masking effect could cause listeners to mistake Vowel 1 for Vowel 2 when the signal is presented along with low frequency BGN.

Nabelek and colleagues ran a series of studies (Nabelek, 1988; Nabelek & Dagenais, 1986; Nabelek, Ovchinnikov, Czyewski, & Crowley, 1996) examining the nature of errors made in vowel perception in background noise. In the two earliest studies (1986, 1988), the authors examined the abilities of normal hearing subjects to identify vowels in the presence of background noise. Nabelek and Dagenais (1986) conducted their study using 10 middle-aged to elderly adults with ranging degrees of sensorineural hearing loss. Fifteen English monophthongs and diphthongs were presented in a quiet condition and in the presence of noise at 0 dB SNR. A monophthong is a single or pure vowel that does not change in resonance throughout the vocalization. These types of vowels can be heard in words like “shoe”, “boat” and “ski”. In contrast, a diphthong is a vowel of changing resonance such as the vowel in words such as “toy”, “how”, & “train” (Borden & Harris, 1984). The stimuli used in the Nabelek and Dagenais (1986) study were presented to a group of normal hearing listening before data collection. All stimuli in these conditions were identifiable by the normal hearing group, but the group with hearing loss had a more difficult time with stimulus recognition.

In 1988, Nabelek followed up her previous study to examine the difference in identification of monophthongs and diphthongs in the presence of background noise for subjects with different degrees of hearing loss. The researchers used the same 15 vowel stimuli as the previous study and presented the stimuli to subjects with ranging hearing levels. This study included a group with normal hearing as well as groups with different degrees of hearing loss. Stimuli were presented in the same conditions as the 1986 study:

in quiet and in background noise at 0 dB SNR. Results of the study indicated that the most errors in the presence of background noise consisted of confusion between monophthongs and diphthongs being perceived as their root monophthongs. These errors were commonly found for monophthongs and diphthongs with 1st and 2nd formants within close proximity. For example: /a/ (“ah”) and /ʌ/ (“uh”) were commonly confused as their first formants are only around 300 Hz apart. This confusion, as well as the number of errors made, increased as the degree of hearing loss worsened.

In their most recent study, Nabelek et al. (1996) examined the effect of background noise at an SNR of 0 dB on the vowel /ai/. Subjects ranged from normal hearing to hearing impaired. Results of this study indicated that the intensity of F₂ needed to be increased in order for vowels to be identified in background noise. However, if these formants were not amplified, the vowels were easily misidentified. These results support the similar findings of the previous studies by Nabelek and colleagues.

Effects of background noise on neural encoding.

Only a few studies have been conducted to investigate the effects of background noise on the neural encoding of a speech signal. Knowing the effects of background noise on neural encoding can aid in the explanation of the common errors in identifying speech signals, as found in several behavioral studies discussed previously (Nabelek, 1988; Nabelek & Dagenais, 1986; Nabelek et al., 1996). The FFR can be an effective EP to study neural encoding of speech in noise because it has the ability to preserve the temporal characteristics of the speech signal such as the envelope or fundamental frequency (F₀) and the TFS or higher formant information.

Although not the FFR, Henry and Heinz (2012) evaluated neural encoding of stimuli in the presence of background noise. In an experiment using chinchillas as opposed to human subjects, electrophysiological responses were recorded to tones in the presence of background noise at SNRs of -10, -15 and -20 dB. Results showed that as the noise increased, the threshold of the response increased. The noise also affected the frequency specificity of the response. This effect of a “broader tuning curve” would result in reduced temporal coding of the signal when in the presence of background noise.

Anderson, Skoe, Chandrasekaran and Kraus (2010) investigated the effects of background noise on the FFR in 60 children. The stimulus used was the speech syllable /da/ and it was presented at 80 dB SPL in quiet and in the presence of multi-talker babble at a +10 dB SNR. The FFR was compared to behavioral task results of the Hearing in Noise Test (HINT). The children were split into two groups based on their HINT performance: one group who scored in the 50th percentile or higher, and a second group who scored below the 50th percentile. Data analysis of the FFR indicated that all children had delayed neural responses (latencies) when the signal was presented along with background noise. Notably, the group that scored below the 50th percentile in the HINT test had greater delays in latency than the group that scored in the 50th percentile or above.

Overall, these results indicate that the adverse listening condition of background noise can result in poor temporal resolution of the speech signal. As previously mentioned, there is not much literature on the effects of background noise on the FFR in several different SNRs. In the current study, we will be further examining neural encoding of speech in background noise for vowel stimuli in normal hearing listeners.

Reverberation

Reverberation consists of multiple reflections of sound energy from the boundaries of an enclosed space (Emanuel & Letowski, 2009). Reverberation time (RT) is the time needed for a sound pressure to decrease in intensity 1000 times (60 dB) after the sound source has ceased its operation. When speech is presented in a highly reverberant environment, the original signal deriving from the speaker is combined with reflections that are time-delayed, scaled versions of the original. Therefore, the reverberant signal that eventually reaches the listener is a combination of direct and reflected energy. Similar to background noise, reverberation is an adverse listening condition that can affect the temporal components of a speech signal, thereby distorting the signal before it reaches the stage of neural encoding.

Effects of reverberation on speech acoustics.

Unlike background noise, reverberation is not a competing sound that is added to the target signal. Instead, reverberation occurs when reflections of the target signal combine with the original signal resulting in distortion. The interactions between the directed and reflected sound waves occurring in reverberation result in “temporal smearing” of the original speech signal (Assmann & Summerfield, 2003). Temporal smearing occurs when a preceding sound segment of a speech signal overlaps the beginning of a subsequent sound segment (Nabelek & Dagenais, 1986). Reflections of the original sound source fill in the areas of the temporal spectra where no vocalization is present, blurring together the components of the signal. This reflection is less detrimental when a speech signal is constant and unchanging in its temporal structure. However,

speech is commonly characterized by rapidly changing spectra. When a reflected and delayed copy blends with the original signal, the relationship of temporal events is blurred (Assmann & Summerfield, 2003). Therefore, reverberation may have little effect on the intelligibility of a pure vowel with no change in resonance, such as a monophthong, but a more detrimental effect on a vowel with a changing resonance, such as a diphthong.

Self-masking is the internal temporal smearing of energy within a phoneme. Nabelek and colleagues used this term to explain the confusion of vowels within a larger speech signal introduced to reverberation. Self-masking can cause a degradation of the F_1 and F_2 transitions in a speech stimulus (Nabelek et al., 1995). Masking of an auditory signal by reverberation depends on the amount of reverberant energy in a room and the amount of decay. The greater the amount of reverberation, and therefore the longer the reverberant decay, the greater masking of the signal that will occur (Nabelek & Dagenais, 1986). This effect of temporal smearing and self-masking is demonstrated in figure 32 below.

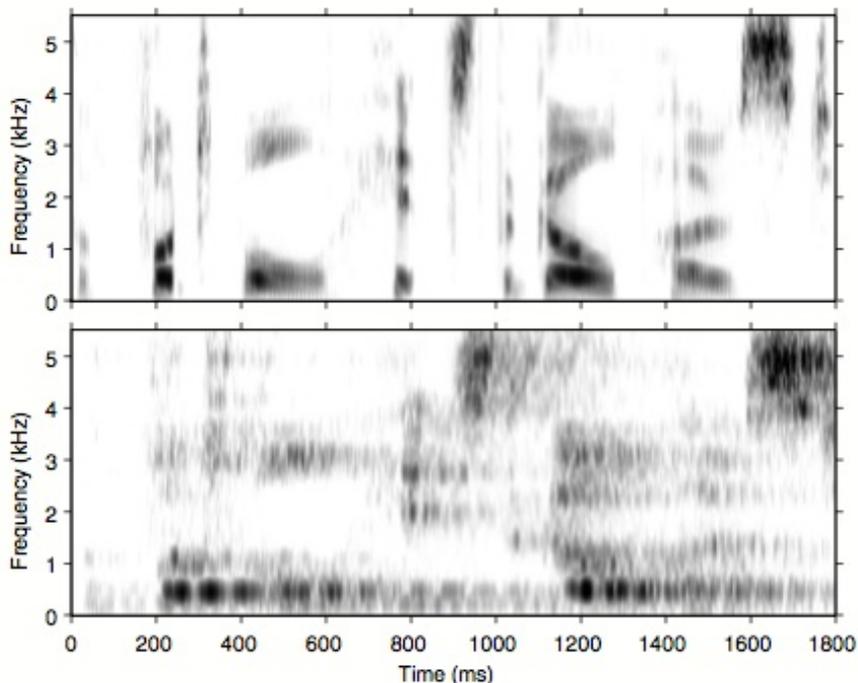


Figure 32. From (Assmann & Summerfield, 2003).

In Figure 32, the top panel shows a speech signal produced in quiet. There are gaps during the vocalization when airflow is stopped (e.g. ~350 ms and 1300 ms). The bottom panel displays the speech signal in the presence of reverberation. Here, the reverberation fills the gaps along the spectra where there is no vocalization (e.g. ~350 ms and 1300 ms). Furthermore, the onsets and offsets of the signal are blurred and the vocalization segments are extended in duration.

Similar to background noise, there are several parameters of reverberation that affect speech intelligibility: ambient noise level, the speaker's vocal output level, the distance between the speaker and the listener, and RT (Assmann & Summerfield, 2003). The current study focuses on the effect of RT on neural encoding of speech. Kreisman (2003) indicated that RT does not begin to affect speech identification in the normal hearing listener until it reaches 1.0 second. However, in listeners with hearing loss, RT

can begin to affect intelligibility of a speech signal in as little as 0.4-0.5 seconds (Kreisman, 2003).

Effects of reverberation on speech understanding.

According to evidence in the literature, studies (Kreisman, 2003; Nabelek et al., 1995) have shown that in environments with moderate levels of reverberation, normal hearing listeners perform fairly well. This is due to the fact that, in general, the formant structures are still the same even if the frequency components have blurred together (Assmann & Summerfield, 2003). In characterizing the effect of reverberation on vowel intelligibility in noise, several studies have summarized the typical errors in vowel production after gathering data on normal hearing and hearing-impaired listeners. As previously mentioned, Nabelek et al. (1995) indicated that the self-masking phenomenon caused by reverberation affected the F_1 and F_2 transitions in the speech signal causing a confusion between monophthongs and diphthongs. In the same study examining the effects of background noise on vowel intelligibility of normal hearing and hearing impaired listeners, Nabelek et al. (1995) also examined the effects of reverberation on vowel intelligibility in these same subjects. Results of this study indicated that with a RT of 1.5 seconds, the 2nd formant of the signal required a greater intensity relative to the other formants in order for the vowel to be correctly identified by listeners. Overall, 90% of the normal hearing listeners were able to identify the vowel correctly. This indicates that although reverberation affects the temporal components of the speech signal, it is still able to be identified the majority of the time by normal hearing listeners, provided that the intensity of the second formant is loud enough.

Nabelek, Letowski & Tucker (1989) also conducted a study to examine the effects of self-masking and overlap-masking on consonant identification. Earlier it was established that self-masking is internal temporal masking within a signal. Overlap-masking, or temporal smearing, is due to the spectral energy of a preceding consonant overlapping or smearing the spectral energy of the following consonant. Eight consonants were presented to the listeners in quiet as well as with a RT of 1.9 seconds. The most common errors made by listeners in the reverberant condition were mistaking /p/ for /t/ and /n/ for /m/. Because these errors were not apparent in the noise conditions, researchers concluded that the errors were due to overlap-masking and self-masking of the signal.

Knowledge of the effect of reverberation on vowel perception is a useful tool when investigating the problems of speech intelligibility as caused by hearing loss. To support this knowledge, it is important to understand the effect of reverberation on neural encoding of speech.

Effects of reverberation on brainstem FFR/neural encoding of speech.

Only a few studies have examined the effects of reverberation on the FFR. Bidelman and Krishnan (2010) examined the effects of reverberation on neural encoding of vowels in 10 musician and 10 non-musician subjects. The study involved the presentation of a synthetic vowel /i/ with a time-varying F_0 ranging from 103-130 Hz and sustained formant frequencies. The stimulus was presented at 80 dB SPL in quiet and in three different RTs: mild (0.7 s), moderate (0.8 s), and severe (0.9 s). Results of the study indicated several findings. First, the temporal resolution of the harmonics decreased

as reverberation levels increased. Second, the F_0 of the signal was preserved in all RT conditions. Third, the F_0 and F_1 components of the response were more pronounced/preserved for the subjects who were musicians than compared to non-musicians. Thus, the smearing of formants rather than the fundamental of the stimulus supports the phenomenon of temporal smearing and self-masking. The errors made in the behavioral studies support these findings. Although this is good information provided by Bidelman and Krishnan (2010), more studies should be conducted on normal hearing listeners to further investigate the effect of reverberation on neural encoding of speech. The current study examines the effect of RT on the FFR to vowel stimuli.

Differences in the Effects of Background Noise and Rreverberation on Speech

Acoustics and Perception.

Both background noise and reverberation affect the acoustics, perception, and neural encoding of vowel stimuli, however, the ways in which each adverse listening condition affects the speech signal differs considerably. Table 2 below summarizes these differences.

Table 2. *Differences between the effects of background noise and reverberation on the speech signal*

	Effects on Speech Signal		
	Acoustics	Perception	Neural Encoding
Background Noise (BGN)	<ul style="list-style-type: none"> • Reduced spectral contrast (peak-to-valley ratio) of entire signal • Masking of higher formant frequency information (F_1-F_3) 	<ul style="list-style-type: none"> • Problems differentiating between monophthongs with F_1 and F_2 formants in close proximity • Interpretation of diphthongs being perceived as their root monophthongs 	<ul style="list-style-type: none"> • Delayed latencies of the FFR in the presence of BGN • Latencies increase with increased severity of BGN
Reverberation	<ul style="list-style-type: none"> • Temporal smearing of vowel • Self-masking of vowel • Degradation of primarily F_2 • F_0 slightly affected, but not as affected as higher formants 	<ul style="list-style-type: none"> • Problems differentiating between monophthongs and diphthongs individually • Problems differentiating between monophthongs with similar F_1 frequencies • Problems differentiating between diphthongs with similar F_1 frequencies 	<ul style="list-style-type: none"> • F_0 and F_1 generally maintained regardless of severity of reverberation • Distortion of higher frequency formants (F_2)

Objectives of the current study.

While numerous behavioral studies have compared the effects of reverberation and BGN on speech acoustics and perception, there are no neurophysiologic studies that have examined the neural encoding of speech stimuli under these two adverse listening conditions. The objective of the current study is to facilitate such a comparison by measuring and contrasting the brainstem FFR in normal-hearing listeners under conditions of BGN and reverberation. Specifically, the two main goals of the current study are as follows:

- Determine the effect of background noise on subcortical envelope and TFS encoding as reflected by the FFR
- Determine the effect of reverberation on subcortical envelope and TFS encoding as reflected by the FFR

Chapter 3:

Methods

Participants

Six normal hearing adults between the ages of 20-35 years ($M=24.5$, $S.D.=0.55$, males=3, females=3) participated in the present study. All subjects had normal hearing sensitivity and normal middle ear function. For the purposes of this study, normal hearing sensitivity was defined as air conduction thresholds of 15 dB HL or better at all test frequencies from 250-8000 Hz. Normal middle ear function was defined as static compliance between 0.3-1.4 mmho, peak pressure between -150 and +100 daPa, and ear canal volume between 0.6-1.5 cc (Katz, 2009). Each participant recruited for this study participated in two 2-3 hour recording sessions at the Towson University Hearing and Balance Center on two sequential test days. Informed consent was obtained from all subjects prior to testing.

Stimulus

A steady-state synthetic vowel /u/ with an $F_0=120$ Hz and $F_1=360$ Hz was generated using the Klatt cascade synthesizer as implemented in Praat (computer generated phoneme producer), as used in Bidelman & Krishnan (2010), was presented to each subject at an intensity of 75 dB SPL. The stimulus was presented as an alternating polarity stimulus at a rate of 3.1/s. During the first recording session, the speech stimulus was presented to the participant in four different background noise (BGN) conditions: a clean condition (no background noise), +5 dB SNR (signal= 75 dB SPL, noise= 70 dB SPL), 0 dB SNR (signal= 75 dB SPL, noise= 75 dB SPL), and -5 dB SNR (signal= 75 dB

SPL, noise= 80 dB SPL). Both the signal and the noise were presented monaurally to the right ear, simultaneously. This was done using two different channels (1 & 2) available in the IHS system. The noise used was speech-shaped noise, which is noise that follows the outline of the speech signal (Emanuel & Letowski, 2009). During the second session, the speech stimulus was presented to each subject in four different reverberant conditions: a clean condition (no reverberation), mild reverberation (RT=0.6 sec), moderate reverberation (RT=0.8 sec), and severe reverberation (RT=1.1 sec). RTs were determined based on use in previous studies examining similar effects (Bidelman & Krishnan, 2010) and were created using MATLAB. Unlike the BGN condition, which uses two different channels to deliver two separate stimuli, the reverberation was added to the stimulus itself and delivered through a single channel.

Procedures

Data was collected using a 2-channel vertical montage using the IHS software. A ground electrode was placed at the lower forehead (Fpz), an active electrode at the high forehead (Fz), two reference electrodes linked at both mastoids (M1&M2), and one reference electrode at the nape of the neck (C7). The stimulus was presented monaurally to the right ear of each subject using electronically shielded ER-3 insert earphones. Inter-electrode impedances were evaluated at each session and maintained below 3Ω . The analog band-pass filter setting was set to 30-3000 Hz. A sampling rate of at least 6,000 Hz was used and 2000 sweeps will be run for each stimulus polarity. The total duration of the averaging window was set to 300 ms and the artifact rejection criteria was set to cut-off at +/- 20 to +/- 75 mV.

Testing took place over two different recording sessions. Each recording session lasted between 2.5 and 3 hours. Stimulus presentation order was randomized in order to eliminate any order effects. Participants were instructed to keep their eyes closed for the duration of the recording, and avoid any unnecessary movements in order to reduce artifacts.

FFR Data Analysis

The FFR was recorded to rarefaction and condensation stimuli. The waveforms recorded to the two stimulus polarities were then added together to provide a summated envelope waveform. The added waveform, after transfer to the frequency domain and evaluated using a Fast Fourier Transform (FFT) technique, yielded information related to the fundamental frequency F_0 of the stimulus. Subtraction of the condensation waveform from the rarefaction waveform provided a TFS waveform. FFT analysis of the TFS waveform provided information on the higher formant frequency (especially F_1) information present in the response.

The resulting data was analyzed using both qualitative and quantitative measures. Qualitative analysis refers to the FFR temporal waveform in clean, BGN, and reverberant conditions. The mean amplitude of the temporal waveform as well as the periodicity of the temporal waveform was analyzed across all test conditions. A spectral analysis was evaluated using MATLAB. Spectral analysis was evaluated within test conditions (a comparison of the clean condition to the different degrees of BGN; a comparison of the clean condition to the different severities of RT). Analysis focused on the bands of

energy present at each frequency of the response as compared to the energy at the frequency bands of the stimulus.

Quantitative analysis was evaluated by use of a Fast Fourier Transform (FFT) analysis technique. As discussed earlier, the FFR is a complex waveform that can be broken down into its respective frequency components. Performing an FFT analysis using MATLAB allowed identification of which frequencies in the response spectrum contain the maximum energy. As the FFR locks on to the stimulus frequencies, maximum energy was present at the fundamental frequency (F_0) as well as the first formant (F_1). Magnitudes of the energy at the F_0 and F_1 frequency regions were obtained for each test subject in each test condition. Due to the limited number of subjects participating in the study, statistical analysis was restricted to descriptive statistics. Descriptive statistics included the mean and standard deviation (SD) of RMS amplitude of F_0 and F_1 for each subject in each stimulus condition. Comparisons of effects of each adverse listening condition were made using tables as seen in the example tables below.

	BGN Condition					Reverberation Condition				
	Clean	+5 SNR	0 SNR	-5 SNR	Difference	Clean	Mild	Moderate	Severe	Difference
Subject 1										
Subject 2										
Subject 3										
Subject 4										
Subject 5										
Subject 6										
M										
SD										

Similar tables were created in order to compare the mean and standard deviation of amplitude of the F_0 and F_1 information for the reverberant condition. A comparison was made between the clean condition and mild, moderate, and severe RTs. Trends were identified across all stimulus conditions as well as within each type of stimulus condition. Differences between effects of these conditions were calculated as well.

Chapter 4:

Results

The following sections of the results will compare and contrast the effects of background noise test conditions and reverberation test conditions on brainstem neural encoding of speech sounds. The results will be split into two major sections: 1) examining the effects of background noise and reverberation on envelope encoding, as reflected by the FFR_{ENV} , which corresponds to the fundamental frequency of the stimulus; 2) examining the effects of the same two listening conditions on TFS encoding, as reflected by the FFR_{TFS} , which corresponds to the first formant of the stimulus. Within each of these primary sections, results from temporal (grand averaged temporal waveforms) and spectral analyses (FFT analysis of F_0 and F_1 , and grand averaged FFT data) will be discussed for both the background noise and reverberation conditions. This same organizational structure will also be employed in the discussion section.

FFR_{ENV} Responses Related to the Fundamental Frequency (F_0)

This section will discuss the results of the added or summed waveform response (FFR_{ENV}), which represents the fundamental frequency of the stimulus (or envelope encoding).

Temporal waveforms- background noise condition.

Figure 33 demonstrates grand averaged time waveforms across six test participants for each background noise condition: clean, +5 dB SNR, 0 dB SNR and -5 dB SNR. Panel A represents the temporal waveform of the FFR_{ENV} in the clean

condition, with no background noise present. Panel B represents the temporal waveform of the FFR_{ENV} in the +5 dB SNR condition. Panel C represents the temporal waveform of the FFR_{ENV} in the 0 dB SNR condition. Panel D represents the temporal waveform of the FFR_{ENV} in the -5 dB SNR condition. This organization of the panels will be followed in all subsequent figures related to the background noise test conditions.

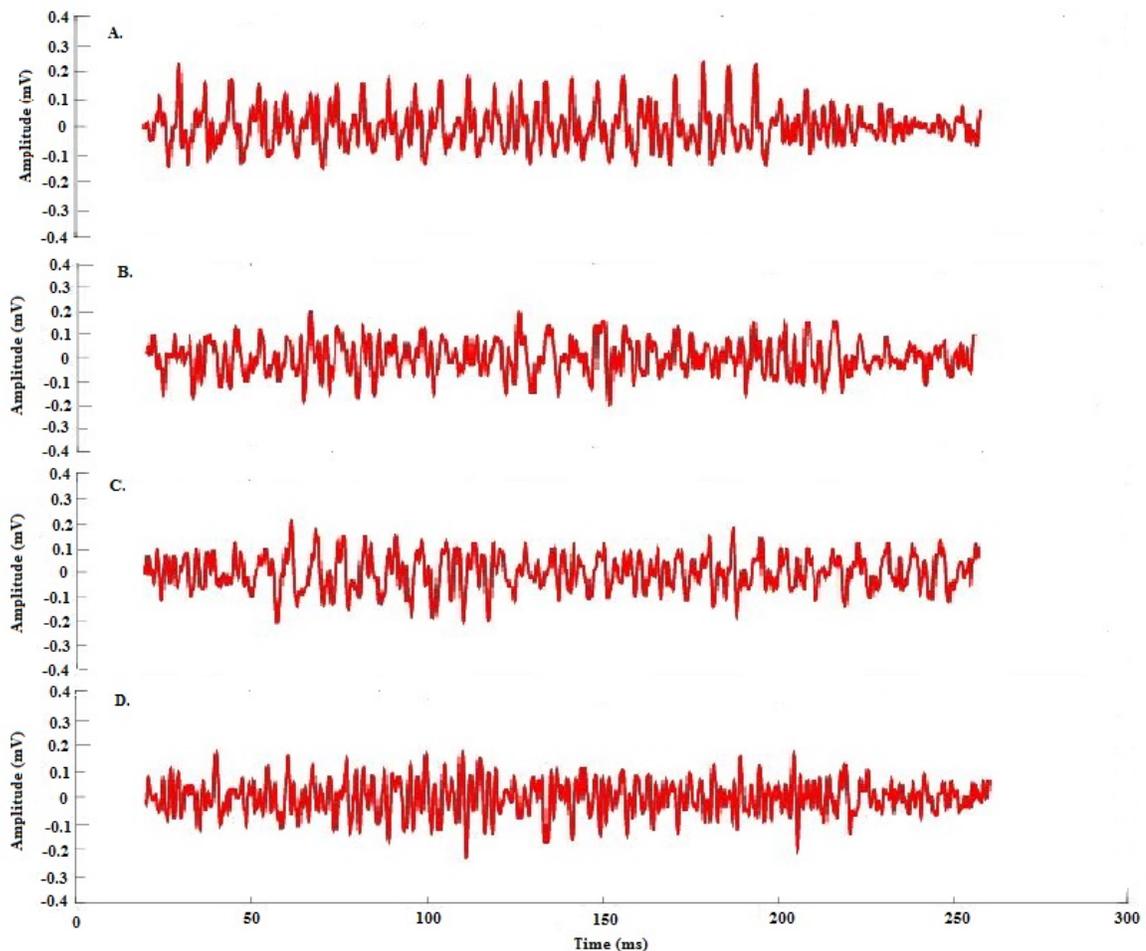


Figure 33. Grand averaged temporal waveforms for varying degrees of background noise (Panel A: clean condition, Panel B: +5 dB SNR, Panel C: 0 dB SNR, Panel D: -5 dB SNR).

Several interesting trends were seen through visual inspection of these time waveforms in both their mean amplitude values as well as the periodicity of the grand averaged waveforms across the four listening conditions. First, as the SNR worsens from

clean to the -5 dB SNR condition, the general amplitude of the response appears to decrease. Note that as the SNR worsens from clean to -5 dB SNR, there is about a 50% reduction in overall amplitude of the FFR_{ENV} response. Specifically, the mean amplitude of the temporal waveform was approximately 0.2 mV in the clean condition and decreased to approximately 0.1 mV in both the 0 dB SNR and -5 dB SNR conditions. Second, the temporal resolution or the periodicity of the waveform becomes less clear and defined as the listening condition worsens. Specifically, note how the periodicity of the response in the clean condition is evident for the entirety of the stimulus presentation (approximately 250 ms). The response pattern is repeatable and well defined with equal amplitude strength out to 250 ms. As the signal degrades with a decrease in SNR, the periodicity of the response worsens so that it is less of a defined periodic pattern. This is especially true in the -5 dB SNR condition.

Temporal waveforms- reverberation condition.

Figure 34 demonstrates grand averaged time waveforms across six test participants for each reverberation condition: clean, mild reverberation, moderate reverberation and severe reverberation. Panel A represents the temporal waveform of the FFR_{ENV} in the clean condition, with no reverberation present. Panel B represents the temporal waveform of the FFR_{ENV} in the mild reverberation condition. Panel C represents the temporal waveform of the FFR_{ENV} in the moderate reverberation condition. Panel D represents the temporal waveform of the FFR_{ENV} in the severe reverberation condition. This organization of the panels will be followed in all subsequent figures related to the reverberation test conditions.

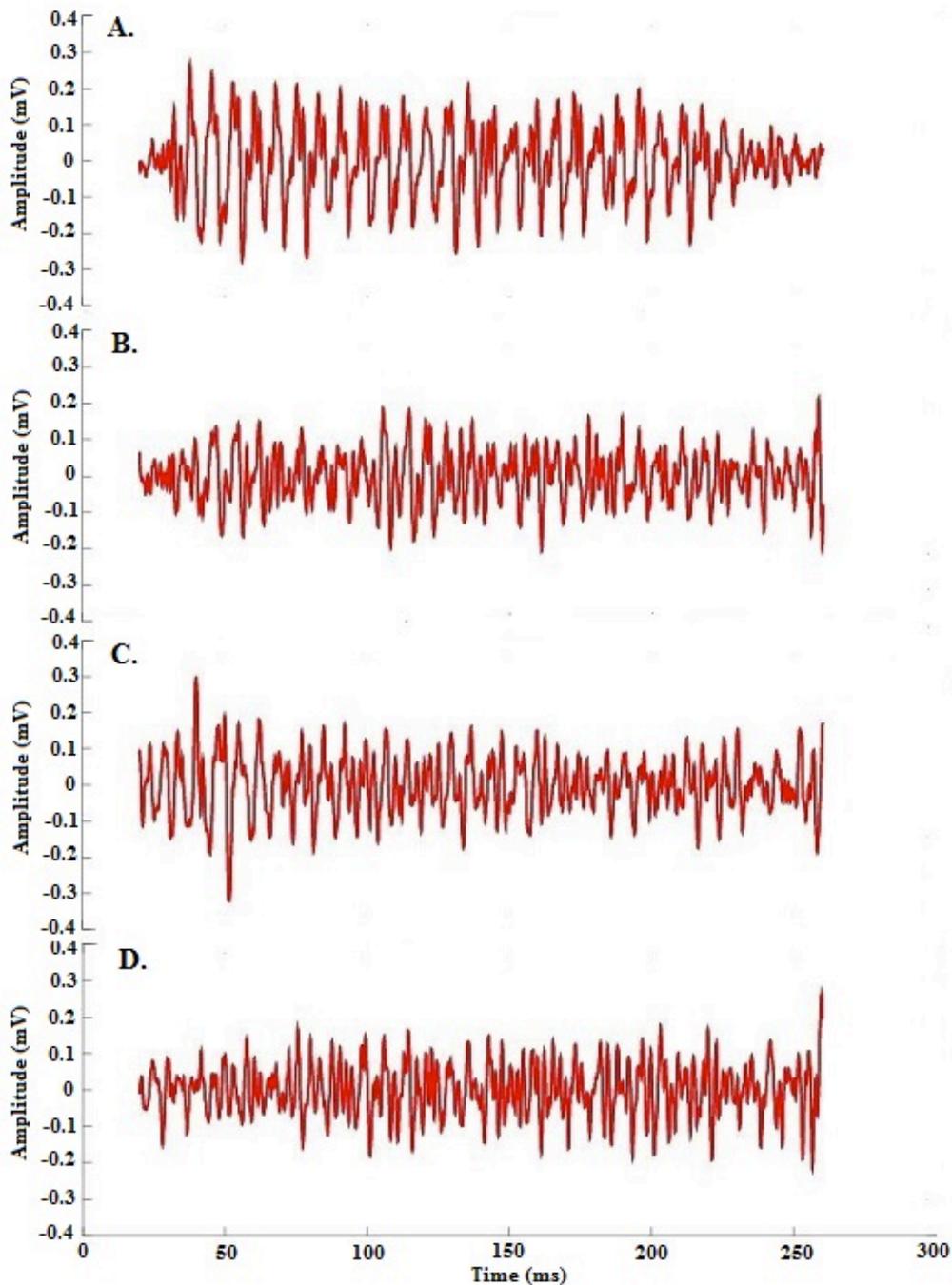


Figure 34. Grand averaged temporal waveforms for varying degrees of reverberation (Panel A: clean condition, Panel B: mild reverberation, Panel C: moderate reverberation, Panel D: severe reverberation).

Similar to the patterns seen in the background noise condition, several interesting trends were noted through visual inspection of the time waveforms in both their overall

mean amplitude values of the responses as well as the periodicity of the time waveforms across the four reverberation conditions. First, as the severity of the reverberation increases, the overall amplitude of the FFR_{ENV} response appears to decrease by approximately a 50% as the severity of the reverberation worsens (as seen from Panels A to D). Specifically, the mean amplitude is approximately 0.2 mV in the clean condition, and decreases to approximately 0.1 mV in the mild, moderate and severe reverberation conditions. A second key finding was that as the reverberation conditions worsen, the periodicity of the response becomes less clear and defined. Specifically, in the clean condition, the periodicity of the response is very clear with a uniform periodicity seen for the duration of the stimulus presentation (~ 250 ms). As the degradation of the stimulus worsens, the temporal pattern of the response becomes less uniform and the periodicity of the response is less evident, especially in the severe reverberation condition. These two patterns of findings are in good agreement with the changes seen in the temporal waveforms for the various SNR test conditions.

Spectral analyses- background noise condition.

FFT analysis was used to convert FFR_{ENV} temporal waveforms into the frequency domain for spectral analysis across both adverse listening conditions. Figure 35 demonstrates averaged FFTs across the four background noise conditions.

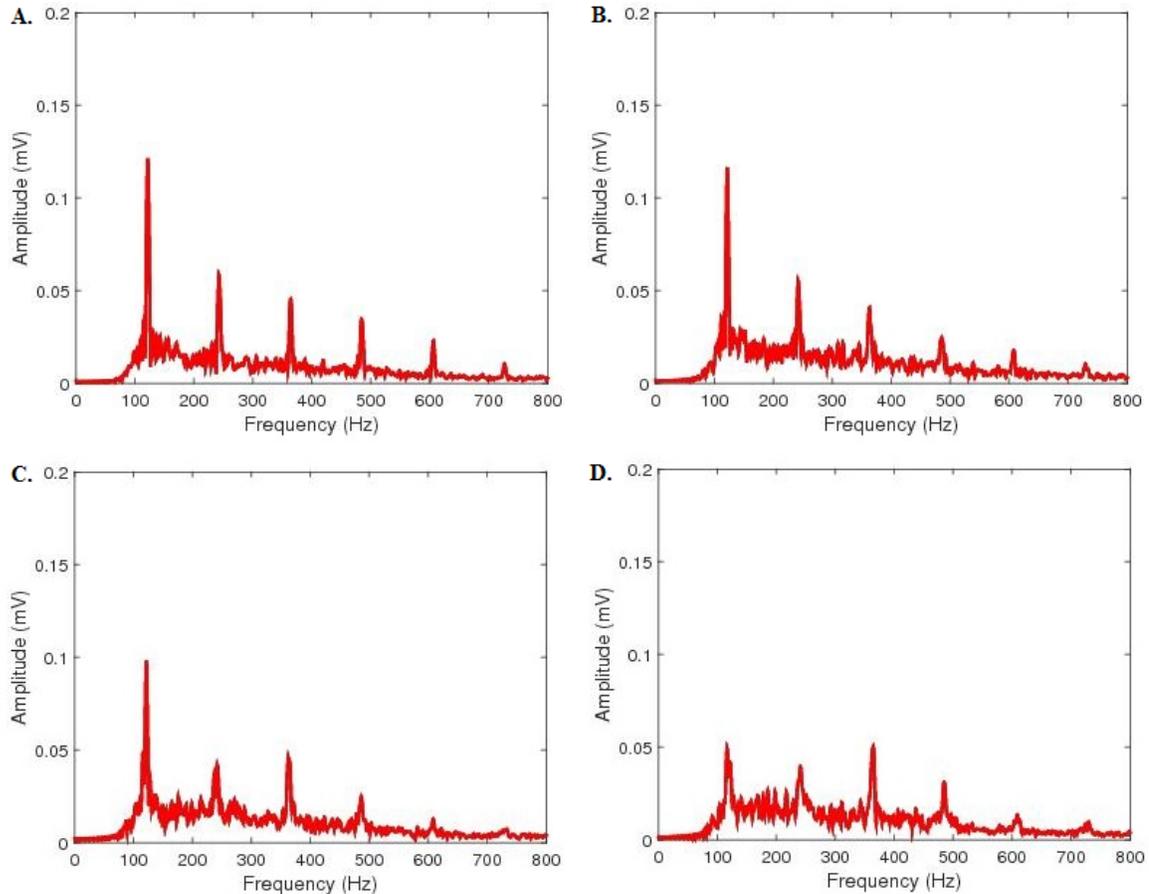


Figure 35. Grand averaged FFTs for varying degrees of background noise (Panel A: clean condition, Panel B: +5 dB SNR, Panel C: 0 dB SNR, Panel D: -5 dB SNR).

Several interesting trends were noted across the FFTs as a function of background noise. First, in the clean listening condition (panel A), there is a clearly defined peak around 120 Hz (F_0) that is visibly larger than any other peaks. As the SNR becomes more unfavorable, the amplitude at the F_0 decreases from approximately 0.12 mV in the clean condition to approximately 0.05 mV in the -5 dB SNR condition. Second, several additional defined peaks occur at approximately whole number multiples of the F_0 as seen in panels A-C. These peaks represent neural encoding at the harmonic frequencies. In each of these panels, the magnitude of energy present at the harmonics changes relative to the F_0 especially for the worse SNR conditions. For example, the magnitude of

energy present at the F0 and the harmonics is essentially the same for the quiet and +5 dB SNR conditions (panels A and B). However, as the signal continues to degrade, the magnitude of energy present at both the F0 and the harmonics decreases, such that in the -5 dB SNR condition (panel D) there is essentially no difference in the energy present at the F0 and its harmonics, indicating a loss of temporal resolution in this condition.

Spectral analyses- reverberation condition.

Figure 36 demonstrates averaged FFTs across the four reverberation conditions.

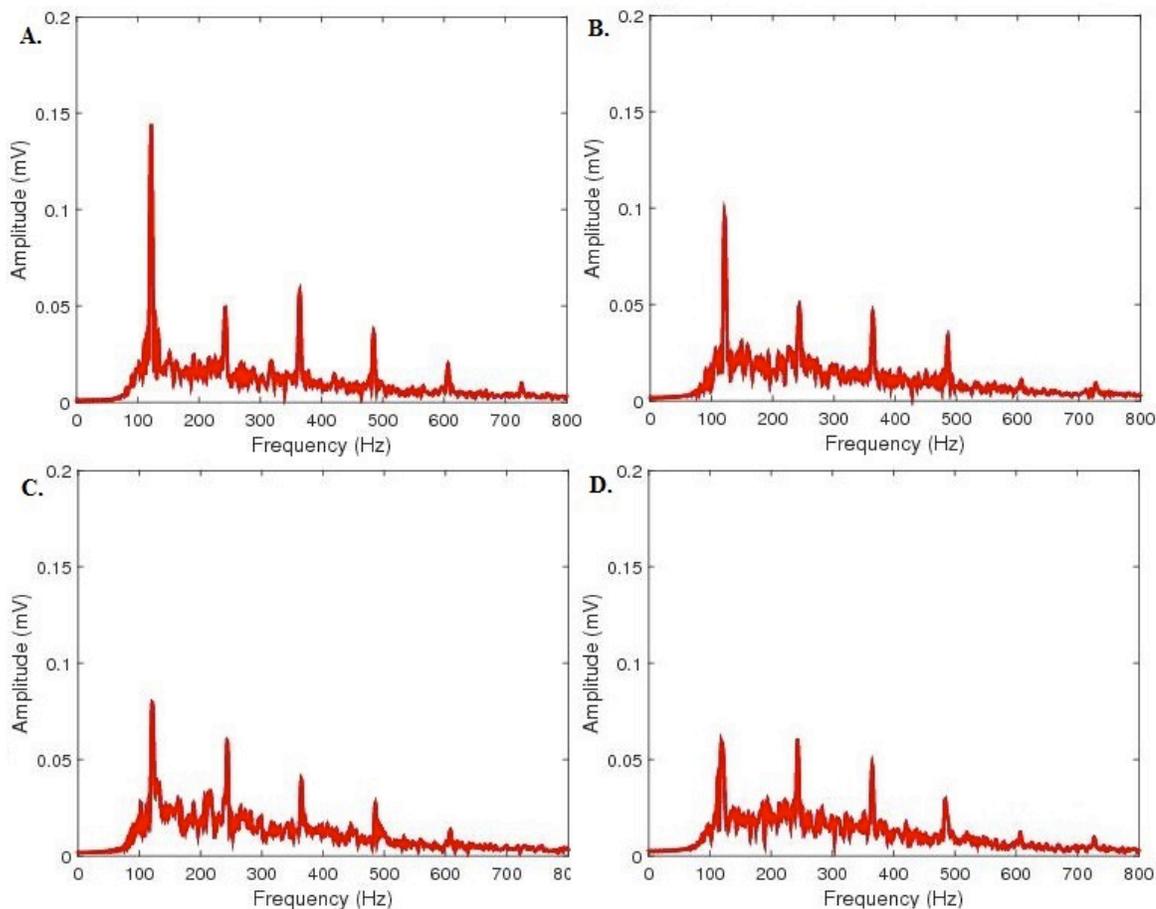


Figure 36. Grand averaged FFTs for varying degrees of reverberation (Panel A: clean condition, Panel B: mild reverberation, Panel C: moderate reverberation, Panel D: severe reverberation).

Several trends were noted across the FFTs in the various reverberation conditions. First: in the clean listening condition, there is a clearly defined peak, which occurs at the F_0 that is visibly larger than any other peaks. In addition, there are also bands of energy at the subsequent harmonic peaks. These spectral findings are in good agreement with those seen in the various background noise test conditions. The two key trends noted in the brainstem neural response obtained in the background noise condition were preserved for that obtained in reverberation. First, as the reverberation condition worsened, the amplitude at the F_0 response decreased. Specifically, the clean condition exhibits the largest amplitude value (.14 mV) and the least favorable reverberant condition (severe reverberation) exhibits the smallest amplitude value (.08 mV). Secondly, the amplitude at the harmonic peaks decreases as the level of reverberation increases. Specifically, the magnitude of energy at the F_0 is always dominant in relation to the level of energy at all the other harmonics in the clean through moderate reverberation conditions (panels A-C). However, the relative difference in amplitude between the F_0 and the harmonics is essentially absent in the severe reverberation condition, thus suggesting that relative harmonic structure is lost in the severe condition. This finding is in good agreement with the -5 dB SNR test condition.

Mean and individual amplitude data at the fundamental frequency (F_0).

In this sub-section, the amplitude values obtained at F_0 for each test subject across the clean and background noise conditions are displayed on the left-hand side of table 3 below. The right-hand side of the table contains individual subject amplitude values obtained at the F_0 across the clean and reverberation conditions.

Table 3. *Individual subjects' amplitude values at the fundamental frequency (F_0) for both background noise and reverberation conditions*

	BGN Condition					Reverberation Condition				
	Clean	+5 SNR	0 SNR	-5 SNR	Difference	Clean	Mild	Moderate	Severe	Difference
Subject 1	0.024mV	0.028 mV	0.025 mV	0.027 mV	-0.003 mV	0.127 mV	0.073 mV	0.076 mV	0.053 mV	0.074 mV
Subject 2	0.151 mV	0.151 mV	0.126 mV	0.025 mV	0.126 mV	0.123 mV	0.163 mV	0.084 mV	0.093 mV	0.03 mV
Subject 3	0.151 mV	0.119 mV	0.109 mV	0.067 mV	0.084 mV	0.119 mV	0.077 mV	0.062 mV	0.063 mV	0.056 mV
Subject 4	0.117 mV	0.175 mV	0.119 mV	0.065 mV	0.052 mV	0.132 mV	0.082 mV	0.081 mV	0.057 mV	0.075 mV
Subject 5	0.115 mV	0.133 mV	0.116 mV	0.065 mV	0.05 mV	0.164 mV	0.072 mV	0.121 mV	0.123 mV	0.041 mV
Subject 6	0.156 mV	0.098 mV	0.108 mV	0.067 mV	0.089 mV	0.182 mV	0.110 mV	0.077 mV	0.065 mV	0.117 mV
M	0.119 mV	0.117 mV	0.100 mV	0.053 mV	0.066 mV	0.141 mV	0.096 mV	0.084 mV	0.076 mV	0.065 mV
SD	0.050 mV	0.051 mV	0.038 mV	0.021 mV	0.029 mV	0.026 mV	0.036 mV	0.020 mV	0.027 mV	0.001 mV

Note. The value listed in the 'Difference' category on the left-hand side of the table (BGN condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the -5 dB SNR condition; The value listed in the 'Difference' category on the right-hand side of the table (reverberation condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the severe reverberation condition.

Mean amplitude values seen in table 3 show a common trend for both the background noise and reverberation conditions. These trends will be discussed individually. First, the background noise amplitude data trends will be discussed followed by the trends found in the reverberation amplitude data.

Background Noise Condition—Amplitude Data Trends

Several interesting trends were noted in the background noise condition (left side of table 3). First, similar to the trends seen visually in the temporal and frequency spectrum figures, the mean amplitude values decrease substantially as the background noise condition worsens. Specifically, the mean amplitude value in the clean condition was 0.119 mV and decreased to 0.053 mV in the most severe listening condition (-5 dB SNR). Second, the variability reflected in the standard deviation values was relatively low across all four test conditions with the range in standard deviations being between 0.021 and 0.051 mV. These low standard deviation values support that the decrease in mean amplitude values across background noise conditions is due to a worsening conditions rather than a wide variability. The mean amplitude values and their standard deviations are displayed in figure 37 below.

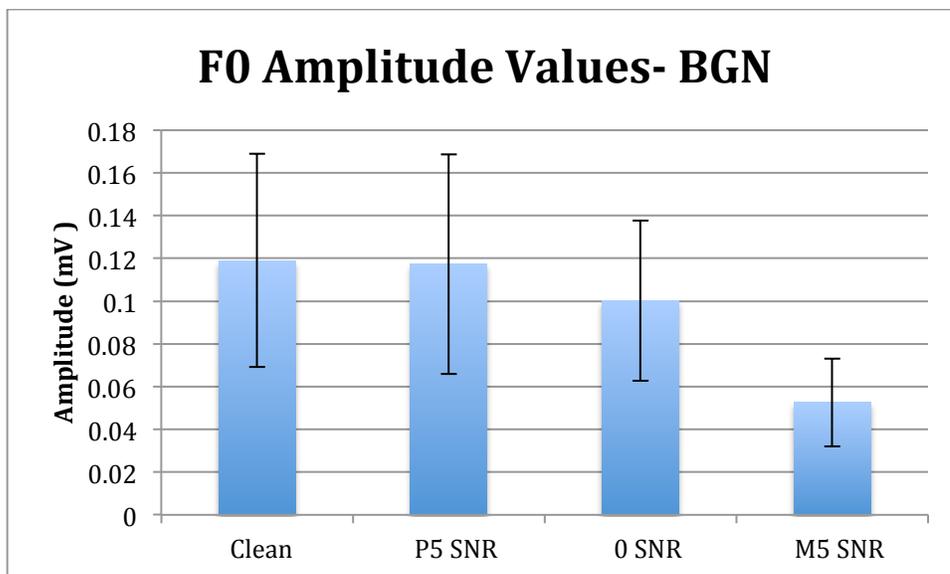


Figure 37. Mean amplitude values for the F₀ across four conditions of background noise.

In general, the trends seen in the mean subject data were also seen at the individual subject level. As the background noise condition worsened, the amplitude at the F₀ decreased. This pattern was true for 5/6 or 83% of subjects and can be seen in the difference values. The only exception to the pattern was seen for subject 1. His/her amplitude values did not change as a function of the background noise conditions.

Reverberation Condition—Amplitude Data Trends

Several trends were noted in the reverberation condition (right-hand side of table 3). First, the mean amplitude values showed a substantial decrease as the reverberation condition worsened. Specifically, the mean amplitude value for the clean condition was 0.141 mV and decreased to 0.076 mV in the severe reverberation condition. Second, the variability in the data, as reflected in the standard deviation values, is essentially similar across the clean versus reverberation conditions (ranging between 0.020—0.036 mV). This supports that the decrease in mean amplitude values seen across conditions is due to

a worsening reverberation condition and not a wide variability. The mean amplitude values and their standard deviations are displayed in figure 41 below.

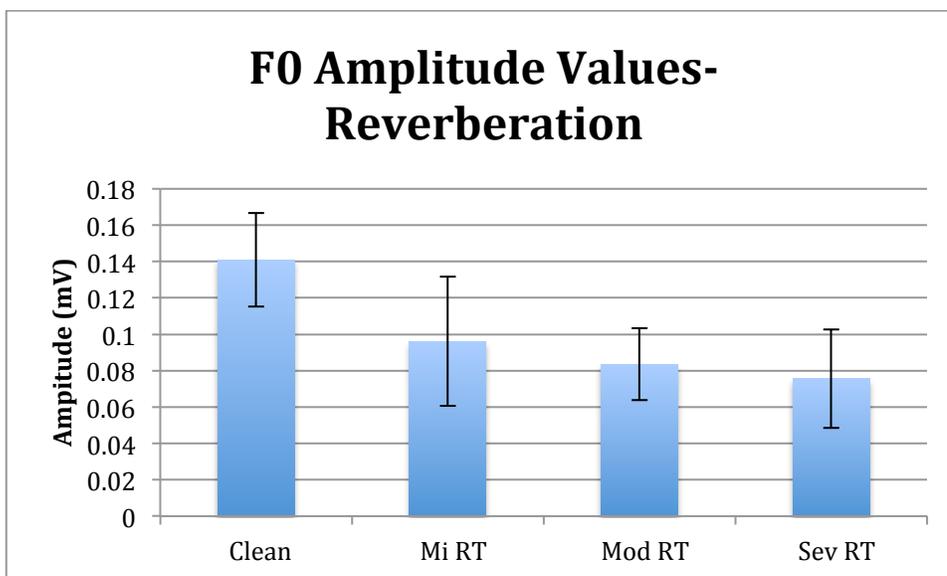


Figure 38. Mean amplitude values for the F0 across four conditions of reverberation.

The individual subjects' amplitude values showed a very similar pattern to the pattern seen with the mean amplitude values across the four reverberation conditions. In general, as the severity of the reverberation increased their F_0 amplitude values decreased. This pattern was true for all six subjects.

In summary, the majority of individual subjects showed a decrease in amplitude at F_0 in going from the clean to worse test conditions (i.e. either the -5 dB SNR or the severe reverberation condition). This pattern was true for both background noise and reverberation conditions. The changes in F_0 amplitude values from clean to most unfavorable condition are seen in the difference values.

FFR_{TFS} Responses Related to the First Formant (F₁)

The following section will discuss the results of the difference or subtracted FFR waveform (FFR_{TFS}), which represents the first formant of the stimulus (or temporal fine structure encoding).

Temporal waveforms- background noise condition.

Figure 39 below demonstrates grand averaged temporal waveforms of the FFR_{TFS} across the four background noise conditions.

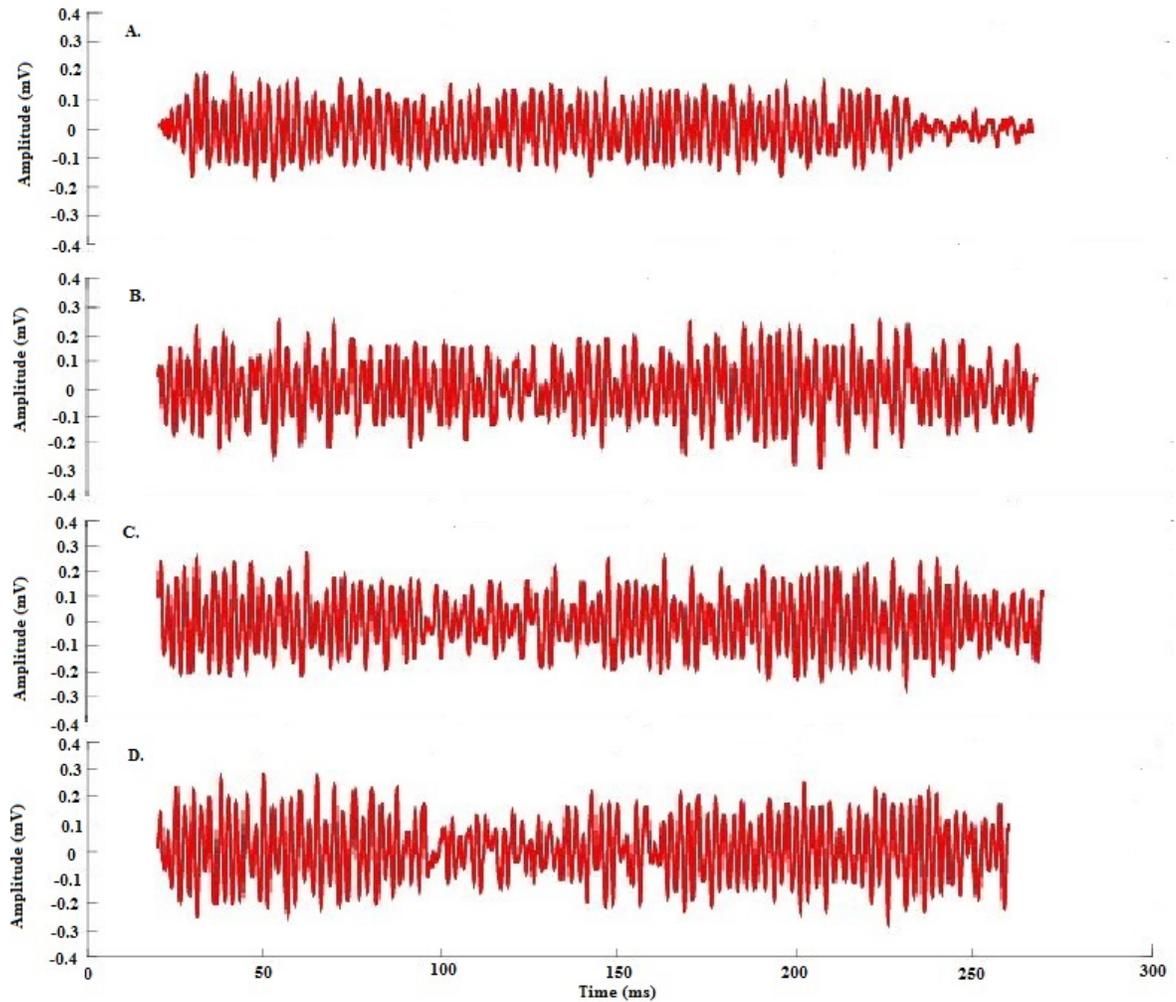


Figure 39. Grand averaged temporal waveforms for varying degrees of background noise (Panel A: clean condition, Panel B: +5 dB SNR, Panel C: 0 dB SNR, Panel D: -5 dB SNR).

A few trends are exhibited through visual inspection of the FFR_{TFS} waveforms in the various background noise conditions. First, there is a slight increase in the amplitude of the F_1 response as the severity of the condition worsens. Specifically, the amplitude of the response in the clean condition is approximately 0.2 mV, and increases to approximately 0.25-0.30 mV across the three background noise conditions. The only exception is seen in the -5 dB SNR condition. The amplitude decreases from approximately 0.25 mV to approximately 0.1-0.15 mV from 100-150 ms and then returns back to approximately 0.25-0.30 mV for the duration of the stimulus presentation.

Temporal waveforms- reverberation condition.

Figure 40 below demonstrates grand averaged temporal waveforms of the FFR_{TFS} across the four reverberation conditions.

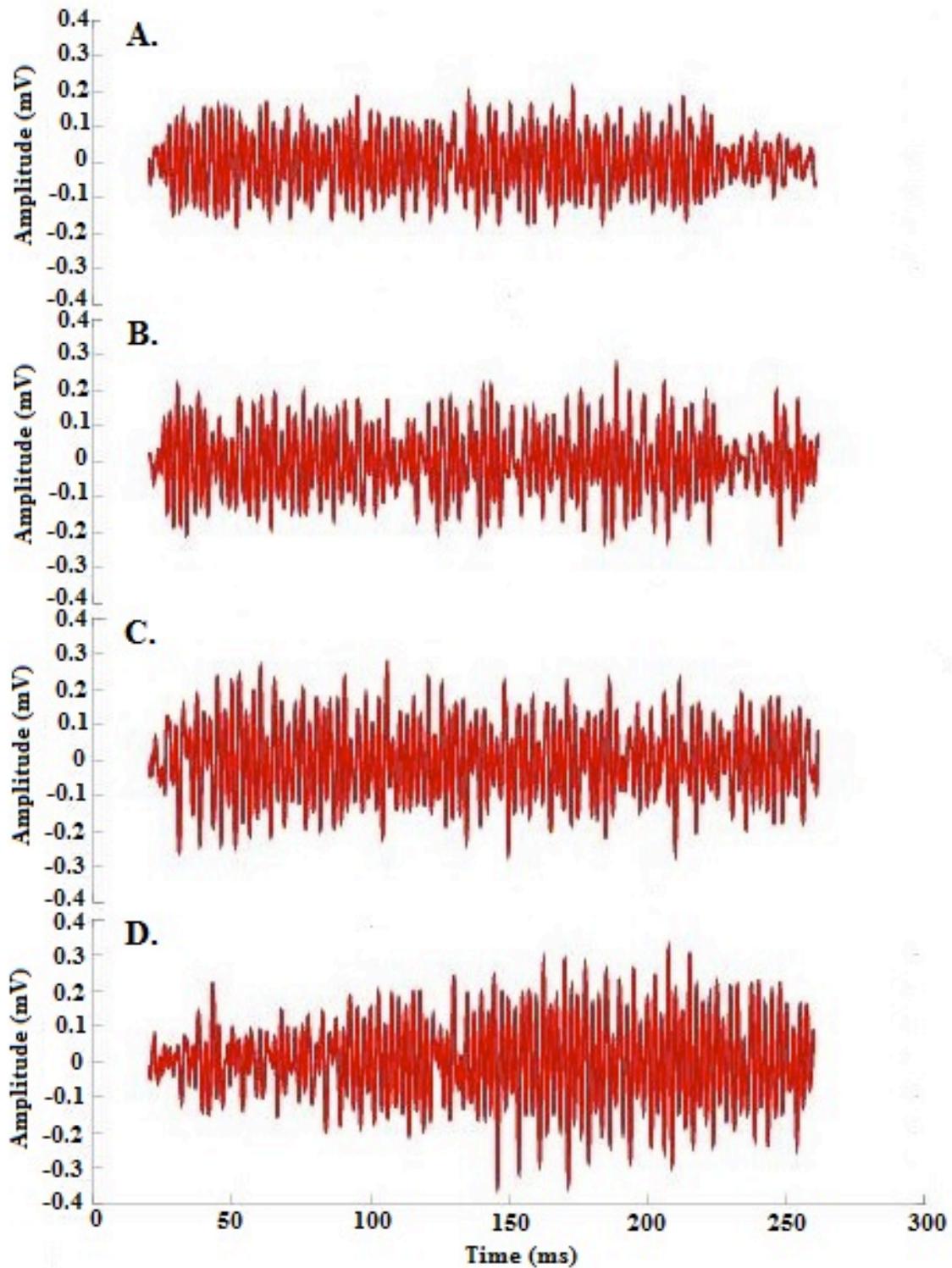


Figure 40. Grand averaged temporal waveforms for varying degrees of reverberation (Panel A: clean condition, Panel B: mild reverberation, Panel C: moderate reverberation, Panel D: severe reverberation).

Several trends were seen in the amplitude of the response as the severity of the reverberation condition worsened. First, as the severity of the reverberation condition worsens the amplitude of the response at F_1 becomes larger. Specifically, the amplitude values in the clean and mild reverberation conditions are approximately 0.2 mV and increase to approximately 0.3-0.4 mV in the moderate and severe reverberation conditions. Additionally, in the severe reverberation condition, the amplitude of the response is approximately 0.1-0.2 mV for the first 140 ms of the stimulus presentation and then increases to approximately 0.3-0.4 mV for the duration of the stimulus presentation.

Spectral analyses- background noise condition.

Figure 41 below demonstrates grand averaged FFTs of the FFR_{TFS} across the four background noise conditions.

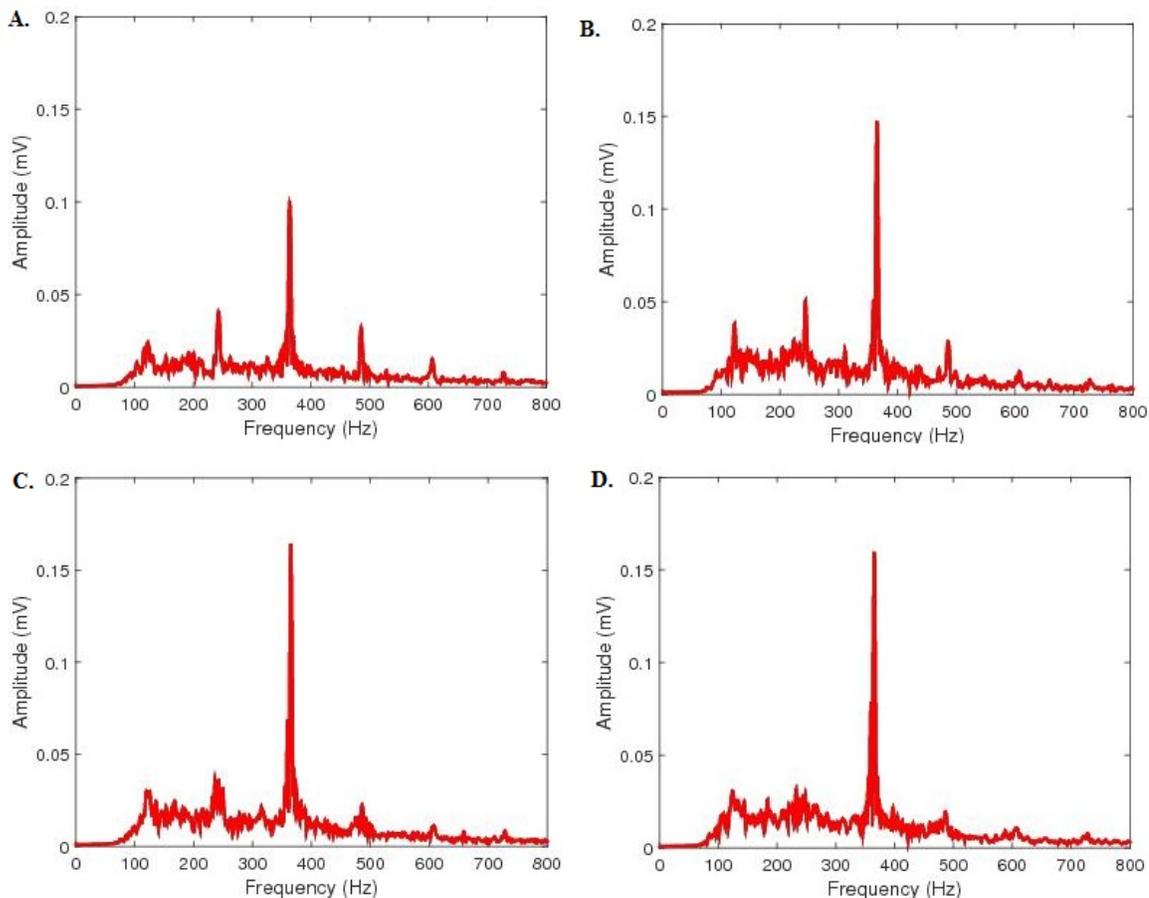


Figure 41. Grand averaged FFTs for varying degrees of background noise (Panel A: clean condition, Panel B: +5 dB SNR, Panel C: 0 dB SNR, Panel D: -5 dB SNR).

Visual inspection across all four conditions shows that as the signal to noise ratio worsened, the amplitude response at F_1 increased. Specifically, the amplitude at F_1 was approximately 0.1 mV in the clean condition and increased to approximately 0.16 mV in the severe reverberation condition. This finding was somewhat unexpected. A second key observation was a change in the magnitude of energy at the harmonics relative to the F_1 . Specifically, in the clean condition, there are visibly defined harmonic peaks surrounding F_1 . However, in the most severe background noise condition (-5 dB SNR), the harmonic peaks are undistinguishable from the surrounding response floor.

Spectral analyses- reverberation condition.

Figure 42 below demonstrates grand averaged FFTs of the FFR_{TFS} across the four reverberation conditions.

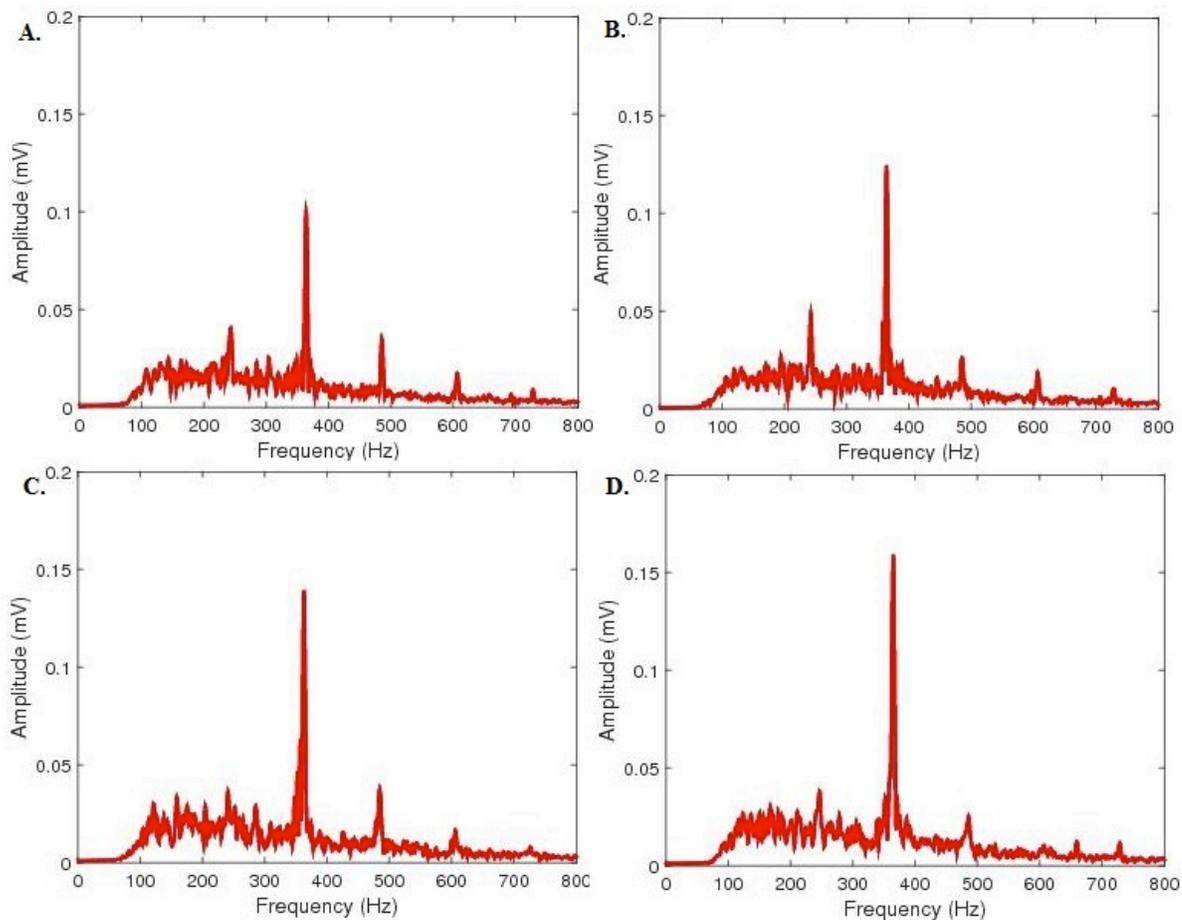


Figure 42. Grand averaged FFTs for varying degrees of reverberation (Panel A: clean condition, Panel B: mild reverberation, Panel C: moderate reverberation, Panel D: severe reverberation).

As seen in figure 45, the amplitude of F_1 *increased* as the degradation of the signal worsened. Specifically, the amplitude at F_1 was approximately 0.1 mV in the clean condition and increased to approximately 0.17 mV in the -5 dB SNR condition. Again, this finding was somewhat unexpected. Secondly, the relative difference between the

harmonic peaks and F_1 was more clearly defined for the less severe conditions, clean and mild reverberation. However, the harmonics are not as prominent in the worst conditions (moderate and severe reverberation).

Mean and individual amplitude data at the first formant (F_1).

The initial analysis of the FFR_{TFS} for both the background noise and reverberation conditions was the same as that conducted for the FFR_{ENV} . However, unexpected findings were obtained. Specifically, for the background noise test conditions, the mean amplitude values at F_1 in the two most severe conditions, (0.15 mV & 0.14 mV, in the 0 and -5 dB SNR conditions, respectively), were larger than the mean amplitudes in the two best background noise test conditions (0.09 mV & 0.13 mV, clean and +5 dB SNR, respectively). Interestingly, the same trend was seen for the reverberation test condition. For example, the mean amplitudes at F_1 in the two most severe conditions (0.13 mV & 0.15 mV, moderate and severe reverberation, respectively) were larger than the mean amplitudes in the two best conditions (0.09 mV & 0.11 mV, clean and mild reverberation, respectively).

In order to investigate the possible explanations for this unexpected response pattern, we examined whether there was a change in the level of the noise floor across these various test conditions that corresponded to the changes in amplitude of F_1 . The noise floor includes both the resting noise that naturally occurs within the body and the additional noise provided by the adverse effects of the degraded stimulus. This investigation revealed that there was in fact a steady increase in the level of the noise floor from the clean to the most severe listening conditions for both the background noise

conditions (clean= .015 mV; -5 dB SNR= .023 mV), as well as the reverberation conditions (clean= .017 mV; severe = .022 mV). This rise in noise floor levels could have led to the increase in amplitude values that was observed as test conditions worsened for both the BGN and reverberation condition.

In order to try and account for this effect, we calculated the derived SNR of the F_1 response in order to normalize the effect of the noise floor level across test conditions. Contrary to the definition of SNR as discussed in the literature review, the term derived SNR used in this section of the paper represents the difference between the sub-cortical response to the stimulus and the overall noise floor. The individual derived SNR values as well as the mean and standard deviation values for both listening conditions were charted in table 4 below. The left-hand side of table 4 contains SNR values obtained at F_1 for each test subject across the clean and background noise conditions. The right-hand side of the table 4 contains individual subjects' amplitude values obtained at F_1 across the clean and reverberation conditions. The trends observed in the F_1 analysis for both the background noise and reverberation test conditions will be discussed individually.

Table 4. Individual subjects' derived SNR values at the first formant (F_1) for both background noise and reverberation conditions

	BGN Condition					Reverberation Condition				
	Clean	+5 SNR	0 SNR	-5 SNR	Difference	Clean	Mild	Moderate	Severe	Difference
Subject 1	1.394 dB	2.054 dB	1.628 dB	1.735 dB	0.341 dB	2.888 dB	3.137 dB	5.962 dB	6.124 dB	3.236 dB
Subject 2	5.425 dB	4.928 dB	5.035 dB	5.040 dB	0.385 dB	3.769 dB	3.485 dB	5.032 dB	4.97 dB	1.201 dB
Subject 3	7.743 dB	6.674 dB	8.782 dB	6.781 dB	0.962 dB	6.559 dB	6.935 dB	6.931 dB	8.083 dB	1.524 dB
Subject 4	5.364 dB	8.031 dB	5.663 dB	6.208 dB	0.844 dB	6.789 dB	6.148 dB	4.762 dB	8.600 dB	1.811 dB
Subject 5	13.189 dB	9.710 dB	9.730 dB	7.451 dB	5.738 dB	8.900 dB	10.509 dB	8.328 dB	9.452 dB	0.552 dB
Subject 6	3.563 dB	9.051 dB	6.069 dB	7.796 dB	4.233 dB	3.394 dB	3.824 dB	5.021 dB	5.346 dB	1.952 dB
M	6.113 dB	6.741 dB	6.151 dB	5.835 dB	0.298 dB	5.383 dB	5.673 dB	6.006 dB	7.100 dB	1.717 dB
SD	4.063 dB	2.867 dB	2.887 dB	2.233 dB	1.83 dB	2.388 dB	2.823 dB	1.394 dB	1.861 dB	0.527 dB

Note. The value listed in the 'Difference' category on the left-hand side of the table (BGN condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the -5 dB SNR condition; The value listed in the 'Difference' category on the right-hand side of the table (reverberation condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the severe reverberation condition.

Table 4. Individual subjects' derived SNR values at the first formant (F_1) for both background noise and reverberation conditions

	BGN Condition					Reverberation Condition				
	Clean	+5 SNR	0 SNR	-5 SNR	Difference	Clean	Mild	Moderate	Severe	Difference
Subject 1	1.394 dB	2.054 dB	1.628 dB	1.735 dB	0.341 dB	2.888 dB	3.137 dB	5.962 dB	6.124 dB	3.236 dB
Subject 2	5.425 dB	4.928 dB	5.035 dB	5.040 dB	0.385 dB	3.769 dB	3.485 dB	5.032 dB	4.97 dB	1.201 dB
Subject 3	7.743 dB	6.674 dB	8.782 dB	6.781 dB	0.962 dB	6.559 dB	6.935 dB	6.931 dB	8.083 dB	1.524 dB
Subject 4	5.364 dB	8.031 dB	5.663 dB	6.208 dB	0.844 dB	6.789 dB	6.148 dB	4.762 dB	8.600 dB	1.811 dB
Subject 5	13.189 dB	9.710 dB	9.730 dB	7.451 dB	5.738 dB	8.900 dB	10.509 dB	8.328 dB	9.452 dB	0.552 dB
Subject 6	3.563 dB	9.051 dB	6.069 dB	7.796 dB	4.233 dB	3.394 dB	3.824 dB	5.021 dB	5.346 dB	1.952 dB
M	6.113 dB	6.741 dB	6.151 dB	5.835 dB	0.298 dB	5.383 dB	5.673 dB	6.006 dB	7.100 dB	1.717 dB
SD	4.063 dB	2.867 dB	2.887 dB	2.233 dB	1.83 dB	2.388 dB	2.823 dB	1.394 dB	1.861 dB	0.527 dB

Note. The value listed in the 'Difference' category on the left-hand side of the table (BGN condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the -5 dB SNR condition; The value listed in the 'Difference' category on the right-hand side of the table (reverberation condition) equals the amplitude value obtained in the clean listening condition subtracted from the amplitude value obtained in the severe reverberation condition.

Background Noise Condition—SNR Data Trends

Several interesting trends were noted for the background noise test condition (left side of table 4). First, as the background noise condition worsened, the mean derived SNR value decreased. For example, the mean derived SNR was 0.74 dB in the +5 dB SNR condition, versus 5.84 dB in the -5 dB SNR condition. Secondly, the variability of the derived SNR values across the background noise conditions was essentially the same in the three background noise conditions. This supports that the decrease in the mean derived SNR values across conditions is truly due to a worsening background noise condition and not due to a wide variability. These derived SNR and standard deviation values are displayed in figure 43 below.

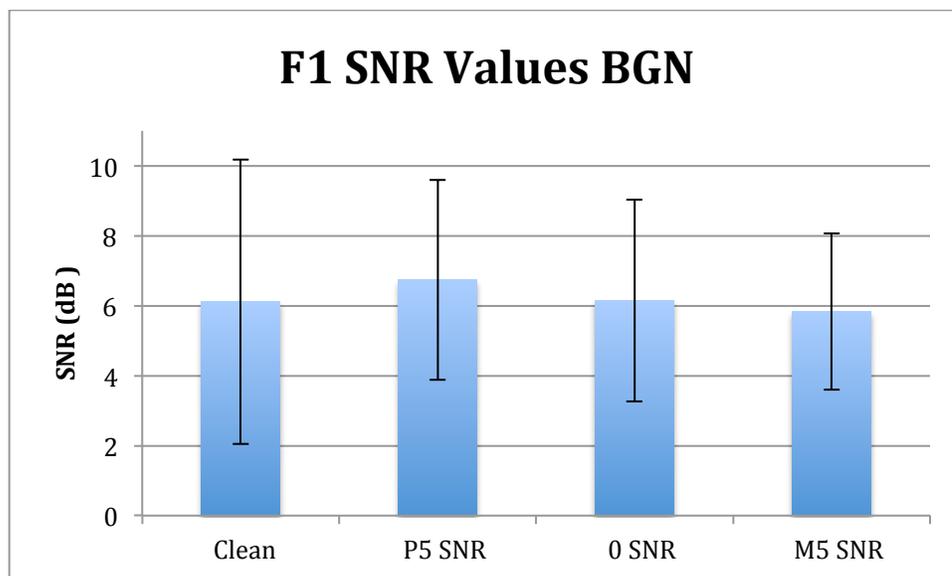


Figure 43. Mean SNR values for the F1 response across four background noise conditions.

In general, the individual data patterns were similar to the pattern as exhibited in the mean derived SNR data. Specifically, 4/6 or 67% of subjects showed a similar decrease in derived SNR value as the background noise condition worsened (+5 dB SNR to -5 dB SNR). The only two exceptions to this were subjects 2 and 3. Both of these

subjects showed essentially no difference in SNR value as a function of background noise condition.

Reverberation Condition—SNR Data Trends

Although F_1 magnitudes in the reverberation conditions were also normalized to the noise floor by using SNR data values, the mean trends of the SNR data continued to be unusual and unexpected for the reverberation condition. Similar to the trend seen while using the absolute amplitude of the response at F_1 , the mean SNR value was the largest for the severe reverberation condition (7.10 dB), and the smallest SNR value was seen in the best or clean condition (5.38 dB). Therefore, the mean SNR values *increased* from the clean condition to the most severe reverberation condition which is the opposite pattern seen in the background noise test conditions. Second, a wider variability range was seen across test conditions (with standard deviation values ranging between 1.39 and 2.82 dB). This change in variability across test conditions may have influenced the unexpected increase in SNR value as the condition worsened. These derived SNR and standard deviation values are displayed in figure 44 below.

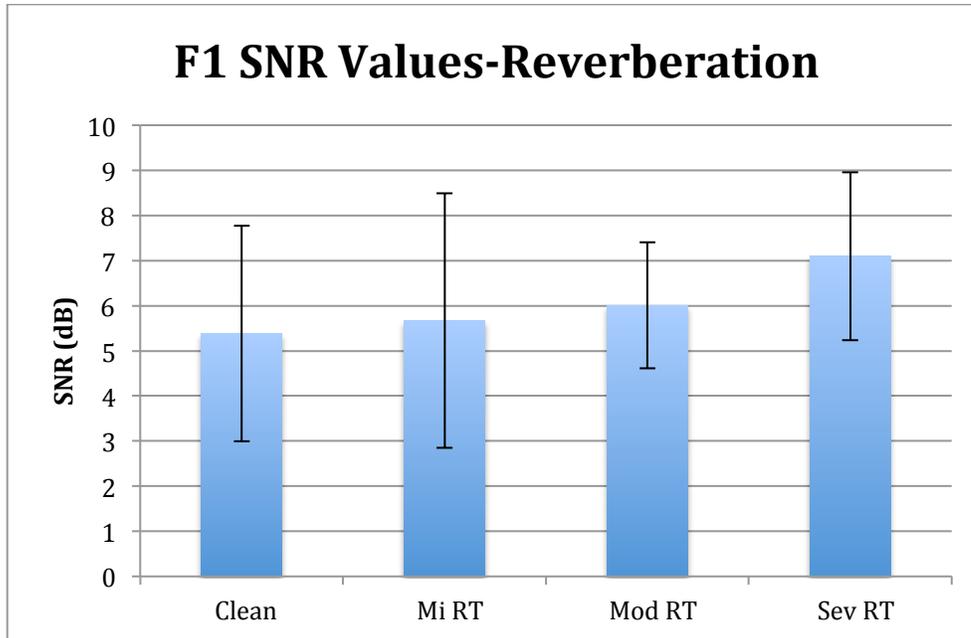


Figure 44. Mean SNR values for the F1 response across four reverberation conditions.

The individual SNR data showed that all six subjects exhibited the same response pattern as the mean data. Specifically, as the severity of the reverberation condition worsened, the SNR response at F_1 increased. The difference between the SNR values in the clean versus severe reverberation condition was as small as 0.552 dB and as large as 3.24 dB. The possible reasons behind these unexpected results will be further investigated in the discussion section of this paper.

The results of the current study are summarized in table 5.

Table 5. Summary of the results for both adverse listening conditions

	Background Noise Condition	Reverberation Condition	Collective Findings
F_0	<ul style="list-style-type: none"> • Decrease in amplitude and periodicity of the temporal waveforms as the severity of condition worsened • Decrease in amplitude at the F_0 as the severity of the condition worsened (seen in FFT and amplitude values) • Relative difference between energy at F_0 and harmonics maintained until the worst BGN condition (-5 dB SNR) 	<ul style="list-style-type: none"> • Decrease in periodicity of the temporal waveforms as the severity of the condition worsened • Decrease in amplitude at F_0 as the severity of the condition worsened • Relative difference between energy at F_0 and harmonics maintained until the worst reverberation condition (severe reverberation) 	<ul style="list-style-type: none"> • Overall degradation of the FFR_{ENV} response with degradation of the signal for both adverse listening conditions
F_1	<ul style="list-style-type: none"> • Increase in overall amplitude of time waveform as the severity of condition worsened • Increase in amplitude at F_1 as the severity of condition worsened (seen in FFT) • Relative difference between energy at F_1 and harmonics maintained until the worst BGN condition (-5 dB SNR) • Decrease in derived SNR values at the worst listening condition (-5 dB SNR) 	<ul style="list-style-type: none"> • Increase in overall amplitude of time waveform as the severity of condition worsened • Increased in amplitude at F_1 as the severity of condition worsened • Relative difference between energy at F_1 and harmonics maintained until two worst reverberation conditions (moderate and severe) • Increase in derived SNR values as the severity of the condition worsened 	<ul style="list-style-type: none"> • Overall degradation of the FFR_{TFS} as seen in the decrease in derived SNR values at the worst listening condition (-5 dB SNR) for background noise condition • Overall increase in amplitude and derived SNR values of FFR_{TFS} with degradation of signal for reverberation condition

Chapter 5:

Discussion

The purpose of the current study was to investigate the effects of two different listening conditions, background noise and reverberation, on the brainstem neural encoding of the speech stimulus /u/ in normal hearing listeners. As previously listed in table 4, there were two main findings of the current study. First, there was an overall degradation of the brainstem envelope encoding or FFR_{ENV} in both background noise as well as reverberation. On the other hand, brainstem encoding of temporal fine structure, or FFR_{TFS} , was affected differently in the background noise and reverberation conditions. Specifically, in the background noise condition, a decrease in FFR_{TFS} was noted as SNR became poorer, while an overall improvement was noted for FFR_{TFS} in reverberation. In order to discuss the findings of the current study, a review of neural encoding at the levels of the basilar membrane and the brainstem is required.

Neural Encoding of Complex Auditory Speech Signals

As the first step of the discussion, we must re-visit the process of neural encoding at the level of the basilar membrane, as this process plays a crucial role in the explanation of our results. Please use figure 45 below as a guide through this topic. As a complex auditory signal approaches the basilar membrane (A), it is separated into its individual harmonic frequencies (B) and subsequently sorted through corresponding auditory filters along the basilar membrane (C). These overlapping band-pass filters are arranged tonotopically (C). That is, low center-frequency filters are at the apex of the cochlea and high center-frequency filters are located at the base. The low frequency filters are narrow

and thus allow fewer frequency components to pass through. In contrast, higher frequency filters are wide and thus allow a broader range of frequencies to pass through. The individual low frequency harmonics/components of the complex signal pass through individual narrow low center frequency bins and result in a simple sinusoidal output also known as a “resolved” harmonic (Bidelman & Krishnan, 2010; Sayles & Winter, 2008). These resolved harmonics are classified as the “temporal fine structure” (D). On the other hand, multiple higher frequency components of the complex signal are passed through individual wider high frequency bins and result in a complex output comprised of several “unresolved” harmonics, which contain both temporal fine structure as well as envelope information. This complex output consists of a slowly-varying envelope superimposed on a rapidly varying fine structure (E). For a more detailed discussion on neural encoding, please refer to page 6 of the literature review.

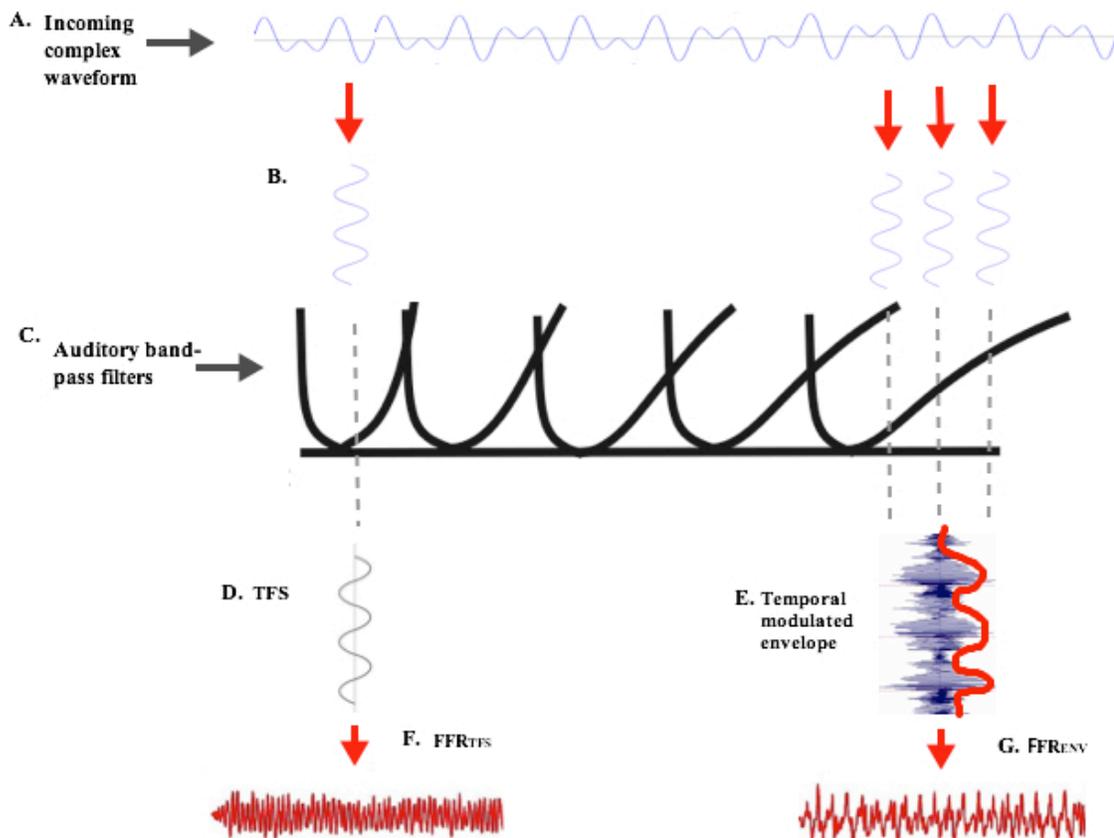


Figure 45. Demonstration of a complex stimulus and its resulting outputs as it passes through various auditory filters along the basilar membrane.

Brainstem Neural Encoding

FFR_{ENV}.

It is important to note that brainstem encoding of *low frequency* F0 or envelope information is primarily mediated by neurons phase locking to the slow-varying envelope at the output of the *high center-frequency filters* (panel E of figure 48). This phase-locking behavior indicates that the FFR_{ENV} is not place-specific (Greenberg, Marsh, Brown & Smith, 1987). This theory was investigated and ultimately supported by researchers Smith, Marsh, Greenberg & Brown (1978) in their study recording the FFR to both a pure tone as well as a complex tone in the presence of masking noise. They found

that when masking noise was added to the pure tone signal, the resulting FFR was severely affected and reduced. The authors conjectured that this was likely due to the place specific neural encoding of the low frequency TFS generated by the pure tone. However, when masking noise centered around the F0 of the complex tone was added, a robust FFR could still be recorded. This finding indicated that neural encoding of low frequency F0 information in a complex tone is mediated by higher frequency regions.

The finding of the Smith et al. (1978) study was upheld in a more recent and unpublished study by Ananthakrishnan (2013), who recorded the FFR to the vowel stimulus /u/ in both normal hearing and hearing impaired adults. When the intact signal was presented to both groups, the normal hearing group yielded more robust FFR_{ENV} responses than that of the hearing impaired group. However, when the stimulus was low pass filtered (i.e. the higher frequency information removed, as typical of SNHL), the normal hearing group showed a decrease in F0 magnitude similar to the hearing impaired group. These findings lend further support to the line of thought that suggests that FFR F0 encoding is highly dependent on high frequency unresolved harmonics rather than the lower resolved harmonics.

FFR_{TFS} .

Harmonic frequencies other than the F0 (lowest frequency) in a complex stimulus are processed through the auditory filters in a place-specific manner as TFS. Lower harmonics are filtered through the low frequency filters as low frequency TFS, which is resolved in nature. On the other hand, higher harmonics are filtered through wider high frequency filters as high frequency TFS, which is unresolved in nature. Hence, TFS

information may be resolved or unresolved, depending on the filter through which it was processed.

The first formant or F_1 related to the speech stimulus /u/ used in the current study is 360 Hz. This is a relatively low frequency TFS which is processed through a narrow low frequency filter to result in a resolved harmonic at the output (as seen in panel D of figure 48). FFR TFS represents phase-locking of brainstem neurons to this resolved TFS (as seen in panel F of figure 48). This behavior is reflected in findings from a second experiment conducted by Ananthakrishnan (2013) where FFR TFS encoding was measured in normal hearing listeners in response to an intact /u/ stimulus as well as a low-pass filtered version designed to mimic high frequency sensorineural hearing loss. Contrary to F0 findings, it was seen that FFR TFS encoding was equivalent in normal hearing listeners for both the intact and low-pass filtered stimuli indicating that brainstem encoding of first formant information is indeed place specific.

Based on the above review, the three main take away points are:

1. Cochlear filter output is resolved for low frequency harmonics and unresolved for high frequency harmonics.
2. FFR envelope encoding is *not* place specific (panel G of figure 48); i.e. it represents brainstem neural encoding to the envelope at the output of the higher center frequency filters that deal with the unresolved harmonics.
3. FFR TFS encoding *is* place specific (panel F of figure 48); i.e. it represents brainstem neural encoding to the higher frequency temporal fine structure information at the output of several different band-pass filters that deal with either resolved or unresolved harmonics.

Effects of BGN & Reverberation on FFR_{ENV}

Background noise and envelope encoding.

The current study found that the FFR_{ENV} degraded as the SNR condition became more unfavorable. This change was noted in both the FFT amplitude data as well as in the temporal waveforms. Further, the response amplitude at F_0 remained relatively robust when the SNR was positive. However, in the worst SNR condition (-5 dB SNR), the amplitude of the response was no longer discernable from the noise floor and its surrounding harmonics. These findings are consistent with similar studies evaluating the FFR_{ENV} in the presence of various background noise conditions (Li & Jeng, 2011; Russo et al., 2004). Both of these studies demonstrated that, as long as the signal-to-noise ratio is favorable, the energy recorded at the F_0 although reduced, is still discernable above the noise floor and subsequent harmonics. However, when the SNR becomes unfavorable (in the current study, this point is reached at -5 dB SNR), there becomes an inability to record a considerable relative difference between the magnitude of energy found at the F_0 and the noise floor.

Similar to the current study, Li and Jeng (2011) used several different SNRs (-12, -6, 0, 6, & 12 dB SNR) to determine at which level the FFR_{ENV} was most affected. Although a reduction in amplitude is noted in the positive SNR condition, the FFR was robust and discernable above the noise floor. They noted that the FFR was “tolerant” of the background noise (and therefore was recordable) until the signal was degraded to the 0 dB SNR condition. However, in the negative SNR conditions, the FFR was not discernable from the noise floor.

Similar results were found in the study conducted by Russo et al. (2004). These researchers recorded the FFR in two positive background noise conditions (+5 and +10 dB) and found that the FFR was robust for both conditions. In order to understand why this decrease in FFR_{ENV} occurs, it is important to first understand how the neural encoding of a speech signal is affected when it is degraded by the addition of background noise. Please use figure 46 as a guide through this topic.

When a complex speech signal is presented in background noise (A/E), the background noise is processed through the cochlear band-pass filters simultaneously with the speech signal (B/F). Hence, the components of the speech signal must compete with background noise at the input and output of every auditory filter. Now, as described earlier, for the speech signal, the low frequency narrow filters (C) yield only the resolved harmonics while the high frequency, wide filters (G) yield unresolved harmonics. However, the level of background noise that passes through these cochlear filters depends on filter-width. Because the lower frequency filters are narrower in width, they allow fewer frequency components of the background noise to pass through (B). As a result, in the case of the low frequency filter output, the level of the resolved harmonics belonging to the target signal are relatively stronger than the level of the background noise. This results in a relatively strong SNR or spectral contrast (D). On the other hand, higher frequency filters are wider, thus allowing more of the background noise energy to pass through (F). Hence, at the output of the higher frequency filters, the level of the unresolved harmonics belonging to the target signal is not as strong compared to the level of background noise. This results in a weaker SNR (H).

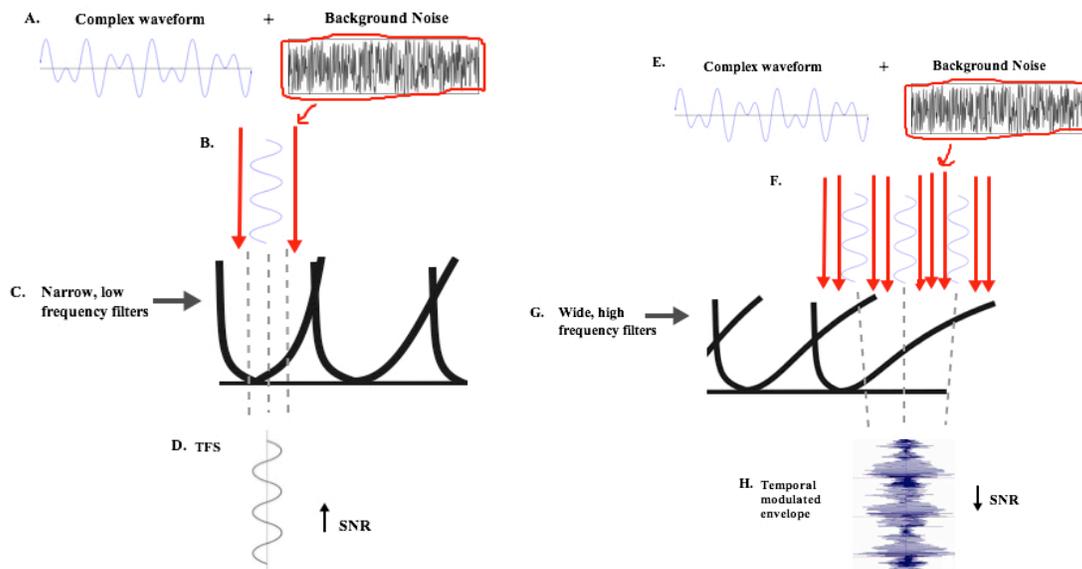


Figure 46. Demonstrates a complex waveform as it passes through the auditory filters along the basilar membrane along with the addition of background noise.

Based on the preceding discussion and the review of neural encoding, we can make the following statements:

- 1) In the presence of background noise, the spectral contrast is significantly reduced at the output of the higher center frequency auditory filters, which deal primarily with the unresolved harmonics (panel D).
- 2) FFR envelope encoding is mediated by neurons phase locking to the output of these same higher center frequency filters dealing with the unresolved harmonics (panel H).

Given these two pieces of evidence, one would expect that in the presence of background noise, a decrease in the F_0 response would be seen as the overall SNR worsens. The results of the current study are consistent with this hypothesis.

Reverberation and Envelope Encoding

The current study found a substantial degradation of the FFR_{ENV} with increasing levels of reverberation, especially in the severe reverberation condition. This reduction

was evident in both the temporal waveforms as well as the amplitude FFT data. These findings are consistent with those seen in the literature. Although limited studies have evaluated the effect of reverberation on F0 encoding, it has been found that increasing reverberation will lead to degraded envelope encoding (Bidelman & Krishnan, 2010; Sayles & Winter, 2008). Bidelman and Krishnan (2010) specifically evaluated the effects of reverberation on the FFR. Similar to the current study, these researchers used three different degrees of reverberation (mild, medium and severe) and found that the FFR_{ENV} was most affected in the most severe condition (with a RT of 0.9s). This degradation of brainstem encoding in reverberation can once again be explained by the neural encoding mechanisms at the level of the cochlea and brainstem.

In contrast to background noise, reverberation is not an additional signal needing to be encoded, but rather a reflected copy of the original signal that leads to a temporal smearing effect seen in the fine structural information (Bidelman & Krishnan, 2010; Sayles & Winter, 2008). This temporal smearing effect as a result of a reverberant environment results in the phase relationships of the harmonics in the signal to be randomized. By the time the complex signal reaches the listener's ear, it possesses a much less modulated temporal envelope than did the original signal when it exited the sound source. Due to the fact that it is the interaction of the unresolved harmonics that creates the TFS and thus the slow-varying envelope, this smearing effect is more likely to affect the higher frequency unresolved harmonics than lower resolved harmonics. Resolved harmonics do not have any other harmonics to interact with at the filter output and therefore are not affected by the opposing phases of other harmonic frequencies (Sayles & Winter, 2008).

The effects of reverberation on envelope encoding can be predicted by the following two statements:

1. Unresolved harmonics are more affected by reverberation than resolved harmonics.
2. FFR TFS encoding is mediated by neurons phase locking to the envelope of the output of wide low frequency filters, which contain these unresolved harmonics.

Given these two points, one would expect a degradation of the envelope response with an increase in reverberation. The results of the current study are consistent with this hypothesis.

Effects of BGN & Reverberation on FFR_{TFS}

Background noise and TFS encoding.

The current study found that when the SNR became the most unfavorable (-5 dB SNR), the FFR_{TFS} was reduced. Further, the magnitude of the energy at the subsequent harmonics was substantially reduced as the SRN worsened. In the -5 dB SRN condition, these harmonics essentially disappear among the noise floor. Few studies have examined the effect of background noise on TFS encoding. Russo et al. (2004) found similar results to the current study. Although they did not record the FFR in negative SNR conditions, they found a significant decrease in F1 magnitude for both positive SNR conditions (+5 and +10 dB). Additionally, although the magnitude at the F1 remained robust, the energy at the higher harmonics was substantially reduced so that the magnitudes at these frequencies were almost completely masked out by the noise floor. Similar to the effect

of BGN on FFR_{ENV} , the effect of BGN on FFR_{TFS} can also be explained by the functioning of the cochlear filter bank and subsequent neural encoding of speech sounds in the presence of background noise.

Leek and Summers (1996) proposed that background noise leads to “broadly tuned cochlear filtering” even in normal hearing adults. This may cause low frequency filters, that are supposed to be narrow, to widen, resulting in the production of *more* unresolved harmonics than would occur in quiet listening conditions. Hence, this broadly tuned cochlear filtering results in an increase in the level of the background noise at the output of even the low frequency filters, with no subsequent increase in the energy at the desired signal. This in turn would lead to further reductions in the spectral contrasts, even at low frequency harmonics that are typically resolved.

The effects of background noise on TFS encoding can be predicted by the following two statements:

1. In the presence of background noise, spectral contrast is reduced even at the output of the lower center frequency auditory filters which deal primarily with resolved harmonics.
2. FFR TFS encoding is mediated by neurons phase locking to the output of these same lower center frequency filters dealing with the resolved harmonics.

Given these two points, one would expect to see a reduction in the amplitude of the energy at F1 as the SNR becomes unfavorable. The findings of the current study are consistent with this hypothesis.

Reverberation and TFS encoding.

The current study yielded unexpected findings for TFS encoding in the presence of reverberation. Specifically, results showed that the magnitude of the energy at the F1 *increased* as reverberation increased. This finding is not consistent with the study as conducted by Bidelman and Krishnan (2010), whose results showed a decrease in F1 encoding as reverberation increased. As mentioned in the previous discussion focused on the effects of reverberation on FFR envelope encoding, it has been established that reverberation has a greater effect on the unresolved harmonics in comparison to the resolved harmonics (Sayles & Winters, 2008; Bidelman & Krishnan, 2010). In the current study, the stimulus used contains an F1 at a low frequency, which yields a resolved harmonic at the output of the cochlear filter-bank. We have also established that FFR_{TFS}, representing brainstem neural encoding of harmonic information, is place specific. Therefore the effects of reverberation can be predicted by the following two statements:

1. Resolved harmonics are less affected by reverberation than unresolved harmonics.
2. FFR TFS encoding is mediated by neurons phase locking to the output of these same lower center frequency filters dealing with the resolved harmonics.

Given these findings and that the F1 of the stimulus used in the current study yields an output of a resolved harmonic, one might predict that reverberation would not have as severe of an effect on the FFR_{TFS} as seen in the response at F₁, as it would on the FFR_{ENV}. The findings of the current study are not consistent with this hypothesis. Although we did not find a degradation of the F₁, we would not expect to see an increase in the magnitude of energy at F1 as reverberation increased.

A reason behind this finding could be due to a low sample size and a high level of variability between conditions. With a larger sample size, variability such as the amount seen in each test condition would have a lesser effect on the mean trends in the data. However, in a small sample size, increased variability may have a greater effect on the mean data trends. For example: subjects 3, 4 & 6 had substantially greater SNR values across conditions than the other three subjects. In these conditions, if even one subject yields extreme results, the mean trends are affected substantially.

Clinical Implications

There are a few implications of the current study. First, mostly behavioral studies have been conducted to evaluate speech understanding in adverse listening conditions (Banai et al., 2009; Johnson et al., 2005; Kraus & Nicol, 2005; Lorenzi et al., 2006). However, few studies have been conducted that specifically evaluate the role of neural encoding on speech understanding in these adverse listening conditions. The results of the current study aid in bridging the gap between brainstem neural encoding and behavioral studies evaluating speech in adverse listening conditions.

Second, a common problem reported by the hearing impaired population is difficulty understanding speech in adverse listening conditions, such as background noise and/or reverberation. By studying the neural encoding of speech at the brainstem level in these adverse listening conditions in individuals with normal hearing sensitivity, it is hoped that this information will shed some insight in understanding why hearing impaired individuals report these difficulties. It is also hoped that this information can be applied to developments in amplification technology. For example: better speech

processing algorithms and/or noise cancellation techniques can be designed for reverberant environments or environments with high levels of background noise. Lastly, a better understanding of the physiology of the cochlea and brainstem pathways can assist hearing scientists in designing technology to better accommodate the types of listening challenges that individuals with sensorineural hearing loss encounter.

Limitations

There were two main limitations of the current study. First, there was a small sample size used. This limited us to only being able to use descriptive statistics. Second, only one vowel stimulus was used in the current study. Because speech is comprised of many different sound segments (monophthongs, diphthongs, consonant-vowel combinations, etc.), it is important to understand the effects of adverse listening conditions for each different sound. This additional information will allow researchers the ability to generalize the effect of adverse listening conditions to the overall understanding of speech.

Future Research

Future research should investigate the neural encoding of the FFR in both adverse listening conditions using larger sample sizes. This would allow for inferential statistics to be employed. It would also be beneficial for a future study to not only record the FFR in these various adverse listening conditions, but also to behaviorally assess these subjects' speech perception capabilities in these same adverse listening conditions. This would allow researchers to evaluate whether or not there is a correlation between a

decrease in an individual's brainstem encoding abilities of these speech stimuli and their results of behavioral speech in noise/reverberation tasks.

APPENDICES

Appendix A

Informed Consent Form

I, _____, agree to participate in a study entitled “The Difference in the Effects of Background Noise and Reverberation on Subcortical Neural Encoding of Speech Stimuli in Normal Hearing Adults,” which is being conducted by Audiology Doctoral Students Randi Cropper, Donald Guillen and Laura Somers, of the Department of Audiology, Speech-language Pathology, and Deaf Studies, Towson University. The purpose of study is to evaluate auditory neural encoding ability in the presence of background noise and reverberation. It is hoped that the information obtained from this study to help explain why normal hearing individuals experience listening difficulty in different adverse listening conditions.

I understand that I must be 18 years of age or older in order to participate in this study. As a participant, I understand that I will be taking part in two test sessions each lasting approximately 2.5 – 3 hours. During these test sessions. I will be asked to relax and sit comfortably in a recliner while electrophysiological recordings are taken from my scalp using scalp electrodes.

I have been informed that any information obtained in this study will be recorded with a unique code number that will allow Randi Cropper, Donald Guillen, Laura Somers and their faculty sponsors to determine my identity. If the data from this study is used in any future publication or professional presentation, my identity will remain confidential and my name will not be used.

I understand that the risk involved with this research is minimal; inline with risk incurred in daily life, as standard electrophysiologic techniques will be employed and the risk is mitigated by the study design. I also understand that my participation is voluntary, and that I am free to withdraw my consent and discontinue participation in this study at any time. If I do withdraw from this study, this will in no way impact any future services I may receive from the Department of Audiology, Speech Language Pathology and Deaf Studies.

If I have any questions or problems that arise in connection with my participation in this study, I should contact Dr. Saradha Ananthkrishnan, the thesis chair at sananthkrishnan@towson.edu / 410-704-6369 and/or Dr. Deb Gartland (chairperson of university IRB committee) at ospr@towson.edu / 410-704-2236.

_____ (Date) _____ (Signature of Participant)

_____ (Date) _____ (Investigator)

THIS PROJECT HAS BEEN REVIEWED BY THE INSTITUTIONAL REVIEW BOARD FOR THE PROTECTION OF HUMAN PARTICIPANTS AT TOWSON UNIVERSITY.

IRB approval number 15-A054 Date of IRB approval 02/23/2015

Appendix B



APPROVAL NUMBER: 15-A054

To: Randi Cropper
2 Waterway Ct APT 2D
Towson MD 21286

From: Institutional Review Board for the Protection of Human
Subjects Scot McNary, Member

Date: Monday, February 23, 2015

RE: Application for Approval of Research Involving the Use of
Human Participants

Office of Sponsored Programs
& Research

Towson University
8000 York Road
Towson, MD 21252-0001

t. 410 704-2236
f. 410 704-4494

Thank you for submitting an Application for Approval of Research Involving the Use of Human Participants to the Institutional Review Board for the Protection of Human Participants (IRB) at Towson University. The IRB hereby approves your proposal titled:

The difference in the effects of adverse listening conditions on subcortical neural encoding of speech stimuli in normal hearing adults

If you should encounter any new risks, reactions, or injuries while conducting your research, please notify the IRB. Should your research extend beyond one year in duration, or should there be substantive changes in your research protocol, you will need to submit another application for approval at that time.

We wish you every success in your research project. If you have any questions, please call me at (410) 704-2236.

CC: S. Ananthkrishnan; P. Korczak
File



Date: Monday, February 23, 2015

NOTICE OF APPROVAL

TO: Randi Cropper **DEPT:** ASLD

PROJECT TITLE: *The difference in the effects of adverse listening conditions on subcortical neural encoding of speech stimuli in normal hearing adults*

SPONSORING AGENCY: None

APPROVAL NUMBER: 15-A054

The Institutional Review Board for the Protection of Human Participants has approved the project described above. Approval was based on the descriptive material and procedures you submitted for review. Should any changes be made in your procedures, or if you should encounter any new risks, reactions, injuries, or deaths of persons as participants, you must notify the Board.

A consent form: is is not required of each participant

Assent: is is not required of each participant

This protocol was first approved on: 23-Feb-2015

This research will be reviewed every year from the date of first approval.


 Scot McNary, Member

Towson University Institutional Review Board

References

- Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology & Neuro-Otology, 11*(4), 213–32. doi:10.1159/000092589
- Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing research, 245*(1), 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 : Official Journal of the International Federation of Clinical Neurophysiology, 119*(4), 922–33.
- Ananthkrishnan, Saradha, "Neural Encoding of Complex Signals in the Healthy and Impaired Auditory Systems" (2013). Open Access Dissertations. Paper 198. doi:10.1016/j.clinph.2007.12.010
- Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience, 30*(14), 4922-4926.
- Arthur, R. M., Pfeiffer, R. R., Suga, N. (1971). Properties of ‘two-tone inhibition’ in primary auditory neurons. *Journal of Physiology, 212*, 593–609.
- Assmann, P., & Summerfield, Q. (2003). The perception of speech under adverse conditions. In S. Greenberg, W.A. Ainsworth, A.N. Popper and R.R. Fay

(Eds.) *Speech Processing in the Auditory System*.

Volume 14, Springer Handbook of Auditory Research.

Banai, K., Abrams, D., & Kraus, N. (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *International Journal of Audiology*, 46(9), 524–32. doi:10.1080/14992020701383035

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and subcortical auditory function. *Cerebral Cortex*, 19(11), 2699–707.

doi:10.1093/cercor/bhp024

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research*, 1355, 112-125.

Borden, G.J. & Harris, K.S. (1984). *Speech Science Primer*. Baltimore, MD: Williams & Wilkins.

Burkard, R.F., Don, M., & Eggermont J.J. (2007). *Auditory evoked potentials: Basic principles and clinical applications*. Philadelphia, PA: Lippincott Williams & Wilkins.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, 47(2), 236–46.

doi:10.1111/j.1469-8986.2009.00928.x

- Clinard, C.G., Tremblay, K.L., Krishnan, A.R., (2011). Aging alters the perception and physiological representation of frequency: Evidence from human FFR recordings. *Hearing Research*, 264(1-2), 48–55. doi:10.1016/j.heares.2009.11.010.Aging
- Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clinical Neurophysiology*, 112(5), 758-767.
- Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *Biomedical Engineering, IEEE Transactions on*, 52(9), 1614-1618.
- 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 and Biobehavioral Reviews*, 35(10), 2046–57.
doi:10.1016/j.neubiorev.2011.05.008
- Duquesnoy, A.J., & Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis. *Journal of the Acoustical Society of America*, 68(2), 537–544.
- Emanuel, D.C. & Letowski, T. (2009). *Hearing Science*. Baltimore, MD: Lippincott Williams & Wilkins

- Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(4), 321-330.
- Galbraith, G. C., & Arroyo, C. (1993). Selective attention and brainstem frequency-following responses. *Biological psychology*, 37(1), 3-22.
- Galbraith, G. C., & Doan, B. Q. (1995). Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 19(3), 203–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7558987>
- Galbraith, G. C., & Kane, J. M. (1993). Brainstem frequency-following responses and cortical event-related potentials during attention. *Perceptual and Motor Skills*, 76(3c), 1231-1241.
- George, E. L. J., Festen, J. M., & Houtgast, T. (2008). The combined effects of reverberation and nonstationary noise on sentence intelligibility. *The Journal of the Acoustical Society of America*, 124(2), 1269–77. doi:10.1121/1.2945153
- George, E. L. J., Goverts, S. T., Festen, J. M., & Houtgast, T. (2010). Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 53, 1429–1440.

- Greenber, S., Marsh, J.T., Brown, W.S., & Smith, J.C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex. *Hearing Research*, 25(2), 91-114.
- Hall, J.W. (2007). *New Handbook of Auditory Evoked Potentials*, Boston, MA, Pearson Education, Inc.
- Hazrati, O., & Loizou, P. C. (2012). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International Journal of Audiology*, 51(6), 437–43. doi:10.3109/14992027.2012.658972
- Jeng, F. C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C. D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear & Hearing*, 32(6), 699-707.
- Johnson, K. L., Nicol, T. G., & Kraus, N. (2005). Brain stem responses to speech: A biological marker of auditory processing. *Ear & Hearing*, 26(5), 424–434.
- 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 : Official Journal of the International Federation of Clinical Neurophysiology*, 119(11), 2623–35. doi:10.1016/j.clinph.2008.07.277
- Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *The Journal of Neuroscience : The Official Journal*

of the Society for Neuroscience, 28(15), 4000–7. doi:10.1523/JNEUROSCI.0012-08.2008

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical “what” and “where” pathways in the auditory system. *Trends in Neurosciences*, 28(4), 176–81. doi:10.1016/j.tins.2005.02.003

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2010). Experience-induced Malleability in neural encoding of pitch, timbre, and timing: Implications for language and music. *Ann NY Academy of Science*, 1169, 543–557. doi:10.1111/j.1749-6632.2009.04549

Kreisman, B. (2003). Stimulated Reverberation And Speech Perception. *The Hearing Journal*, 56(6), 14-15.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels, *Hearing Research*, 166, 192–201.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, 189(1-2), 1–12. doi:10.1016/S0378-5955(03)00402-7

Krizman, J., Skoe, E., & Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology and Neurotology*, 15(5), 332-342.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience : The Official*

Journal of the Society for Neuroscience, 29(18), 5832–40.

doi:10.1523/JNEUROSCI.6133-08.2009

- Levi, E.C., Folsom, R.C., & Dobie, R.A. (1995). Coherence analysis of envelope-following responses (EFRs) and frequency-following responses (FFRs) in infants and adults. *Hearing Research*, 89, 21-27.
- Li, X., & Jeng, F. C. (2011). Noise tolerance in human frequency-following responses to voice pitch. *The Journal of the Acoustical Society of America*, 129(1), EL21-EL26.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B.C.J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of National Academy of Sciences*, 103(49), 18866-18869.
- Marsh, J. T., Brown, W. S., & Smith, J. C. (1974). Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalography and Clinical Neurophysiology*, 36(4), 415–24. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4140069>
- Marsh, J.T., Worden, F.G., & Smith, J.C.(1970). Auditory frequency-following response: Neural or artifact? *Science*, 169(3951), 1222–1223.
- Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology: JARO*, 9(4), 399–406.
doi:10.1007/s10162-008-0143-x

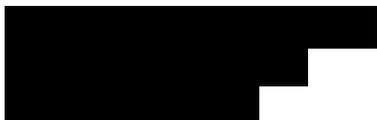
- Moushegian, G., Rupert, a L., & Stillman, R. D. (1973). Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, 35(6), 665–7. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/4128165>
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, 241(1), 34-42.
- Nabelek, A.K. (1988). Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss. *Journal of the Acoustical Society of America*, 84(2), 476-484.
- Nabelek, A. K., & Dagenais, P. A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners, *Journal of the Acoustical Society of America*, 80(3), 741–748.
- Nabelek, A.K., & Nabelek, I.V. (1994). Room Acoustics and Speech Perception. Katz, J. (ed.) from *Handbook of Clinical Audiology*, Baltimore, MD: Williams and Wilkins, pp. 621-637
- Nabelek, A.K., Ovchinnikov, A., Czyzewski, Z., & Crowley, H.J. (1996). Cues for perception of synthetic and natural diphthongs in either noise or reverberation. *Journal of the Acoustical Society of America*, 99(3), 1742–1753.

- Neuman, A. C., Wroblewski, M., Hajicek, J., & Rubinstein, A. (2010). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear & Hearing, 31*(3), 336–344.
- Pickett, J. M. (1957). Perception of vowels heard in noises of various spectra. *The Journal of the Acoustical Society of America, 29*(5), 613-620.
- Picton, T.W. (1990). Auditory evoked potentials. In Daly DD & Pedley Ta (eds.) *Current Practice of Clinical Electroencephalography*, 2nd edition. New York: Raven Press, pp. 625-678.
- Plyler, P. N., & Ananthanarayan, A. K. (2000). 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.
- Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 115*(9), 2021–30.
doi:10.1016/j.clinph.2004.04.003
- Sayles, M., & Winter, I.M. (2008). Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron, 58*(5), 789-801.
- Skoe, E., & Kraus, N. (2011). Auditory brainstem response to complex sounds: a tutorial. *Ear & Hearing, 31*(3), 302–324. doi:10.1097/AUD.0b013e3181cdb272.Auditory

- Smith, J.C., Marsh, J.T., & Brown, W.S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, *39*, (465-472).
- Smith, J.C., Marsh, J.T., Greenberg, S., & Brown, W.S. (1978). Human auditory frequency-following responses to a missing fundamental. *Science*, *201*(4356), 639-641.
- Svirsky, M.A., Teoh, A., & Neuburger, H. (2004). Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. *Audiology & Neuro Otology*, *9*, 224-233.
- Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, *62*(4), 463–469.
doi:10.1016/j.neuron.2009.05.002.Learning
- Wever, E.G., & Bray, C.W. (1930). The nature of acoustic response: The relation between sound frequency and frequency of impulses in the auditory nerve. *Journal of Experimental Psychology*, *13*(5), 373–387.
- Wile, D., & Balaban, E. (2007). An auditory neural correlate suggests a mechanism underlying holistic pitch perception. *Plos One*, *2*(4), e369.
- Worden, F., & Marsh, J. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42–52. doi:10.1016/0013-4694(68)90085-0

CURRICULUM VITA

NAME: Randi Cropper



PROGRAM OF STUDY: Audiology

DEGREE AND DATE TO BE CONFERRED: Doctor of Audiology; May, 2016

Secondary education:

Leonardtown High School
Leonardtown, MD
May, 2008

Collegiate Institutions Attended	Dates	Degree	Date of Degree
Towson University	Fall 2009- May 2012	Bachelor of Science	May 2012
Towson University	Fall 2012-May 2016	Doctor of Audiology	May 2016

Major: Audiology

Professional positions held:

- **Jacksonville Hearing and Balance Institute, LLC:** *Audiology Extern*
10475 Centurion Pkwy N
Jacksonville, FL 32256
- **ENTAA Care:** *Audiology Student Clinician*
203 Hospital Dr.
Glen Burnie, MD 21061
- **Johns Hopkins Listening Center:** *Audiology Student Clinician*
601 North Caroline Street - 6th Floor
Baltimore, MD 21287
- **Maryland School for the Deaf:** *Audiology Student Clinician*
101 Clarke Pl
Frederick, MD 21701
- **Towson University Hearing and Balance Center:** *Graduate Student Clinician*
1 Olympic Pl
Towson, MD 21204