TOWSON UNIVERSITY
OFFICE OF GRADUATE STUDIES

THE EFFECT OF REVERBERATION ON SUBCORTICAL NEURAL ENCODING OF SPEECH STIMULI IN NORMAL-HEARING ADULTS

by

Donald Guillen

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 Donald Guillen B.A, Au.D. Candidate, entitled The Effect of Reverberation on Subcortial 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. Saradh — 5/4/2015

Saradha Ananthakrishnan, Ph.D.
Chair, Thesis Committee — Date

PEGGY KORCZAK (signed by SARADHA ANANTHAKRISHNAN) — 5/14/2015

Peggy Korczak, Ph.D.
Committee Member — Date

Stephen Pallett — 5/13/2015

Stephen Pallett, Au.D.
Committee Member — Date

Janet V Delany — 5-15-15

Janet Delany
Dean of Graduate Studies — Date

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis committee for all their hard work, patience, and support throughout this entire project. Dr. A, thank you for introducing me to this topic and for your countless explanations on this difficult material. Dr. Korczak, thank you for keeping us organized and always encouraging us to keep fighting on when things got tough. I would also like to thank Dr. Paul Evitts, who took his personal time to review and explain fundamental speech acoustics with us. The completion of this project would not have been possible without the support of my loved ones. I would like to especially thank Elaine Edelstein, my family and girlfriend for their countless words of encouragement and prayers. I love you all and this accomplishment is as much yours as it is mine. Lastly, I would like to thank my fellow doctoral students for their continued support and suggestions throughout this entire process.

ABSTRACT

The Effect of Reverberation on Subcortical Neural Encoding of Speech

Stimuli in Normal-Hearing Individuals

Donald Guillen

There is limited objective research on the brainstem representation of speech in degraded listening environments. The aim of this study was to determine the effect of reverberation on brainstem neural encoding in six normal hearing individuals using the Frequency Following Response (FFR). The FFR was recorded in response to a /u/ (F0=120 Hz, F1=360) stimulus in four different levels of reverberation (clean, mild: 0.6 RT, moderate: 0.8 RT, severe: 1.1 RT). FFR data was divided into two types, a qualitative data analysis and a quantitative data analysis. Qualitative indices of the neural response at each RT were provided through a visual analysis of the periodicity and root mean square (RMS) amplitude of the grand averaged waveforms obtained at each reverberation condition. FFT spectral analysis was utilized to measure the amplitudes of both the F0 and F1 components for all participants in each condition. As expected, as reverberation severity was increased, F0 encoding ability decreased. This trend was relatively consistent across participants, especially in the severe reverberation condition. In contrast, there was an unexpected increase in F1 encoding ability with increased reverberation. Results from this study suggests that reverberation, regardless of severity causes a decrease in F0 neural encoding ability. However, reverberation effects are less substantial for F1 encoding. These differences in neural encoding ability between F0 and

F1 components appear to be related to differences in the encoding of these two frequency components at the physiological level of cochlea.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

KEY TO ABBREVIATIONS

SNHL: Sensorineural Hearing Loss

CANS: Central Auditory Nervous System

FFR: Frequency-Following Response

cABR: Complex Auditory Brainstem Response

E: Envelope

TFS: Temporal Fine Structure

CNS: Central Nervous System

PSVEPs: Pattern-shift Visual Evoked Potentials

SEPs: Short-latency Somatosensory Evoked Potentials

ABR: Auditory Brainstem Response

AEPs: Auditory Evoked Potentials

PCPs: Processing Contingent Potentials

F0: Fundamental Frequency

F1: First Formant Frequency

CM: Cochlear Microphonic

CN: Cochlear Nucleus

SOC: Superior Olivary Complex

IC: Inferior Colliculus

MOC: Medial Olivary Complex

SS: Steady-state

TV: Time-varying

AM: Amplitude Modulated

FFT: Fast Fourier Transform

ER-3A: Etymotic Research 3A

IRN: Iterated Rippled Noise

ISI: Interstimulus Interval

SOA: Stimulus Onset Asynchrony

SNR: Signal-to-noise Ratio

LL: Lateral Lemniscus

FDLs: Frequency Discrimination Difference Limens

VCN: Ventral Cochlear Nucleus

PL: Primary-like

RMS: Root Mean Square

FFR_{ENV}: Envelope Encoding

FFR_{TFS}: Temporal Fine Structure Encoding

CHAPTER 1: INTRODUCTION

One of the most common complaints of hearing aid users with sensorineural hearing loss (SNHL) is that “they can hear what people are saying, but cannot understand them.” This is especially true when these individuals are present in degraded or adverse listening environments, such as in the case of background noise and reverberation. In essence, it is a problem of clarity, not audibility (Killion, 1997). With the advent of digital hearing aid circuitry and directional microphone technology, current audiological practices emphasize noise management for clarity. However, even with these advances in hearing aid technology, many hearing aid users still experience significant difficulty in adverse listening conditions. Therefore, it is necessary to look directly at how the normal central auditory nervous system (CANS) encodes complex speech stimuli presented in adverse listening conditions to determine some underlying causes of this problem.

An effective audiological tool that has been used for analyzing auditory encoding abilities at the level of the brainstem in adverse listening conditions is the Frequency Following Response (FFR) or Complex Auditory Brainstem Response (cABR). This electrophysiological response provides a window into the ability of the auditory brainstem to encode speech signals. Specifically, the FFR provides insight into two spectral components of the speech signal, the slowly varying envelope (E) and the temporal fine structure (TFS), which are neural cues necessary for processing and recognizing speech.

Sayles and Winter (2008) have suggested that individuals with SNHL have a decreased ability to accurately code TFS information present in speech signals in reverberant conditions compared to normal hearing individuals. This occurs because

SNHL results in a compromise of the TFS feature of the signal. Moreover, it has previously been reported that individuals with SNHL depend more on E cues for speech perception ability rather than TFS cues (Sayles & Winter, 2008). Given that E cues are also affected by reverberation, the speech perception ability of individuals with SNHL is significantly poorer than those with normal hearing. The effect of reverberation on speech perception is not to be underestimated; previously reported negative effects confirm that more research is needed to understand human auditory neural encoding in this condition. In order to understand the effect of reverberation in individuals with SNHL, it must first be understood in normal hearing participants. The current study will focus on the FFR to reverberant conditions in normal hearing participants; therefore, the following literature review will focus primarily on this topic.

Prior to administering and interpreting the FFR, an audiologist must have a thorough understanding of numerous topics related to the response. These topics include the history of the response; basic speech acoustic terminology; auditory neural encoding at and below the level of the brainstem; optimal recording, stimulus, and subject parameters for recording the FFR; and expected effects of reverberation on the FFR. The organization of the ensuing literature review will emphasize understanding on these various topics.

CHAPTER 2: LITERATURE REVIEW

The human body is capable of processing information from various sensory modalities due to the function of the central nervous system (CNS). There are several branches within the CNS responsible for receiving and processing sensory-specific information. The CNS responds to external stimuli by producing electrical signals called evoked potentials (Chiappa, 1997). The production of an electrical response by the CNS indicates the reception/detection of a stimulus. Evoked potentials are not limited to a single sensory modality within the CNS; potentials can be measured across various sensory organs. These sensory organs include the visual, auditory, and somatosensory modalities. Pattern-shift visual evoked potentials (PSVEPs) and short-latency somatosensory evoked potentials (SEPs) correspond to responses recorded from the visual and somatosensory systems, respectively. In contrast, the auditory brainstem response (ABR) represents an example of neural activity recorded to auditory stimuli from the lower portion of the CANS (Chiappa, 1997). Auditory evoked potentials (AEPs) are electrical signals generated by the CNS in response to auditory stimulation. For some types of AEPs, the electrical response to the signal can indicate either conscious or unconscious discrimination of acoustic information.

Evoked potentials, regardless of sensory modality, have many clinical applications and when used in conjunction with other tests can serve as diagnostic tools to assess these various sensory modalities. According to Chiappa (1997), there are three clinical applications for using evoked potentials: first, to indicate a change or abnormality in the functioning of the sensory system when routine evaluation is not possible; second, to locate or isolate the progression of pathology anatomically; third, to establish a baseline and observe changes in the sensory system over time. There are many types of evoked potentials specific to the auditory system. Picton (1990) proposed a classification scheme to help categorize these AEPs. This classification scheme is described below.

Classification System of AEPs

The first part of Picton's classification scheme refers to the stimulus-response relationship. There are three types of stimulus-response relationships in AEPs. These are transient, sustained, and steady-state responses. Transient responses are elicited by rapid changes in the auditory stimulus. In a transient response, the neurons in the auditory system only fire at the onset or initial phase of the stimulus. On the other hand, sustained responses are potentials where neural firing is maintained for the duration of the stimulus. Lastly, steady-state responses are transient-like responses except that these are provoked by very quick repetitive stimuli that cause multiple neural firings. These responses overlap in the post-stimulus analysis window. As a result, a continuous-like response waveform is produced.

Another classification scheme looks to the temporal characteristics or latency of potentials to describe AEPs. According to Hall (2007), latency is expressed in milliseconds (ms) and refers to the point in time where a response is evoked after a stimulus is presented. This classification scheme can be divided into five subtypes. Picton (1990) defined these subtypes as first (0-1 ms), fast (1-10 ms), middle (10-75 ms), slow (75-300 ms), and late (300+ ms). It has been suggested that different latency values signify neural firing at specific anatomical structures (Atcherson & Stoody, 2012).

The underlying neural generators that are the anatomical sources of potentials have also been used for classifying AEPs. The anatomical sources of auditory evoked potentials corresponding to the latency subtypes (discussed above) are: first (cochlea), fast (the vestibulocochlear nerve and various brainstem structures), middle (high sub-cortical and cortical structures), and slow and late (cortical, higher order structures) (Atcherson & Stoody, 2012).

The last classification scheme differentiates AEPs as either sensory or processing-contingent potentials (PCPs). When a response is dependent on extrinsic events, it is referred to as a sensory (exogenous) potential. On the other hand, responses independent of extrinsic factors that occur due to causes solely *within* a system are referred to as PCPs (endogenous). According to Donchin, Ritter, and McCallum (1978), there are several primary criteria needed to identify a potential as endogenous. First, the potential should not be obligated to respond to a stimulus. Thus, the potential may exist with no stimulation and neural firing can occur when no stimulus is present. Secondly, the amplitude and latency characteristics of the potential should not be related to those of the stimulus used to elicit the response.

Collectively, each of these schemes composes Picton's comprehensive classification of human auditory evoked potentials. Based on Picton's scheme, the FFR is classified as a sensory, fast (2-20 ms), and sustained potential of brainstem origin. Since the FFR will be recorded in the present study, the focus of the remainder of this literature review will be on this response.

Basic Acoustics

In order to both administer and interpret the FFR, clinicians must have a good understanding of basic speech acoustics and auditory neural encoding. The following section will go into detail about how the human auditory system processes sounds and will also introduce basic speech acoustic terminology that relates to the FFR.

There are two types of acoustic waveforms, simple and complex waveforms. A simple waveform is also known as a pure tone since it is composed of a single frequency. This waveform is typically described as being sinusoidal (as seen on left side of Figure 1). The term “sinusoidal” refers to the even repetition of the waveform. On the other hand, a complex waveform is composed of many tones that are either related (periodic) or not related to each other (aperiodic). In a complex waveform, if the frequencies that compose the waveform are harmonically related, then the complex waveform is periodic (as illustrated on right side of Figure 1 below). If the frequencies that compose the waveform are not harmonically related, the complex waveform is aperiodic.

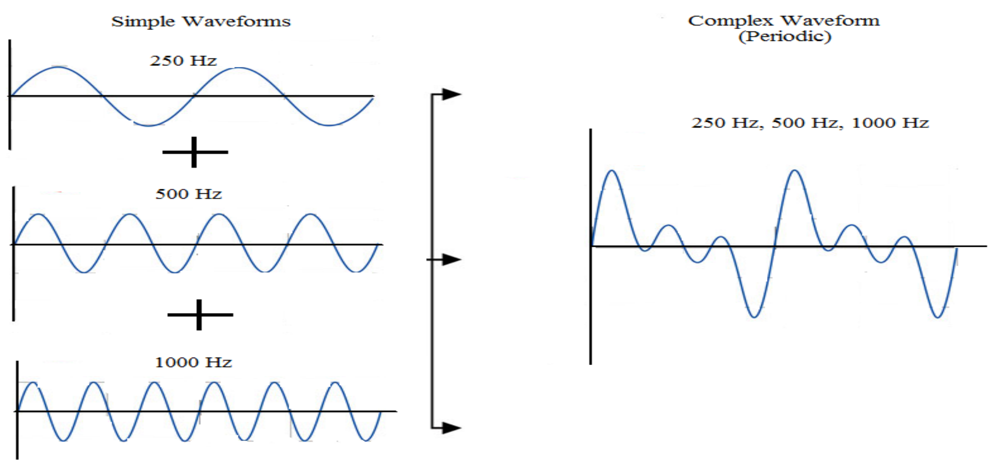

Figure 1. This complex waveform is periodic because it is composed of pure tones that are harmonically related (500 Hz and 1000 Hz are multiples of 250 Hz), all of these waveforms repeat at discrete intervals.

Periodicity. Periodicity refers to discrete repetition in a waveform. A waveform can be either periodic or aperiodic (no repetition of waveform). If a waveform is periodic, then the component frequencies are harmonically related and occur at whole number multiples of one another. An example of a periodic waveform is a pure tone (as seen on left side of Figure 2). If a waveform is aperiodic, then the component frequencies are not harmonically related and each tone occurs at frequencies independent of one another. An example of an aperiodic waveform is noise (as seen on right side of Figure 2).

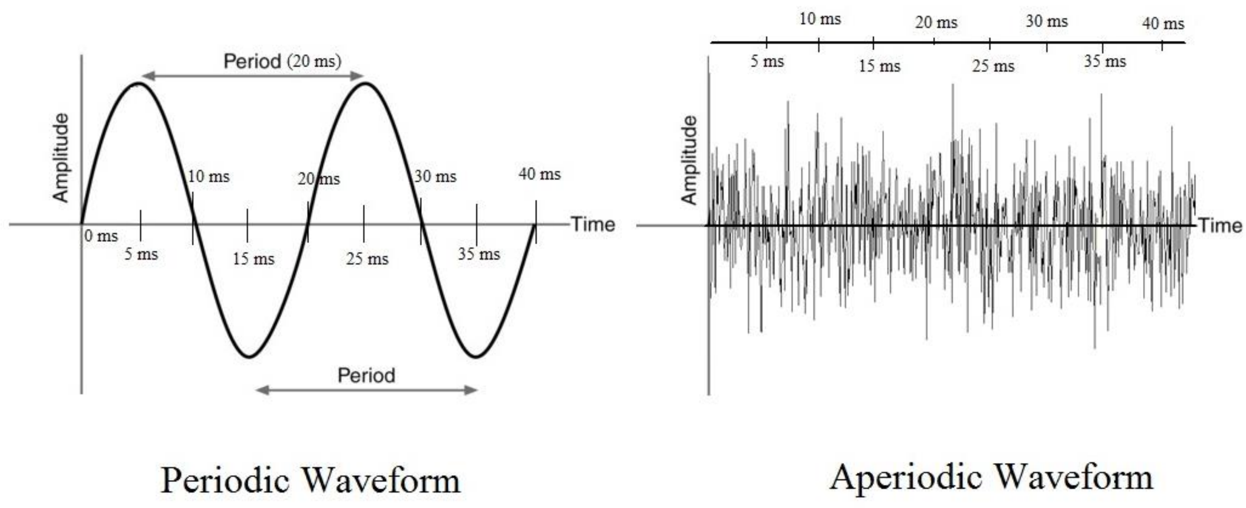

Figure 2. The left side of the figure shows a periodic waveform which illustrates discrete repetition every 20 milliseconds. This illustration shows a simple waveform; however, periodic waveforms can also be complex. The right side of the figure shows an aperiodic waveform with no discrete repetition.

Acoustics of Speech. Speech is a complex stimulus. It is therefore important to understand that there are many components that compose the acoustics of human speech. The underlying component of speech is the fundamental frequency. The fundamental frequency (F0) refers to the lowest frequency of a pattern repetition (Raphael, Borden, & Harris, 2008), or the lowest frequency of a vibrating system with a periodic waveform

(Emanuel & Letowski, 2009). In speech, the F0 is determined by the rate at which the vocal folds vibrate.

Directly related to the F0 are the harmonics of speech. A harmonic is a whole number multiple of the F0. For example, a frequency that is two times the F0 is known as the second harmonic, whereas a frequency that is five times the F0 is known as the fifth harmonic. The harmonics of the speech vowel /u/ can be visualized in the spectrogram below.

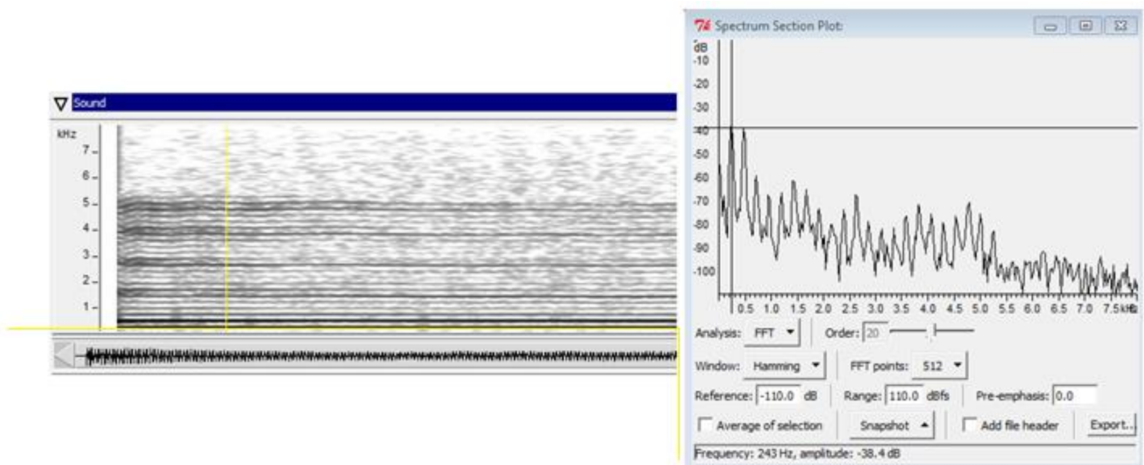

Figure 3. Narrowband spectrogram with better frequency resolution showing the harmonics (left hand side). A Fast Fourier Transform (FFT) of speech syllable /u/ showing the F0 (highest peak at 243 Hz) and harmonic frequencies (right hand side).

In the spectrogram above, each horizontal striation represents a harmonic, that is, complex frequencies created by the source (vocal chords). The FFT on the right side is a representation of those harmonics converted into the frequency domain from the temporal domain. The fundamental frequency of this stimulus is 243 Hz. Each harmonic thereafter is a whole number multiple of 243 Hz.

Superimposed on top of the harmonics are formant frequencies. Formant frequencies are frequencies in the speech signal where there is an increased concentration of acoustic energy. In other words, formant frequencies are acoustic resonances created by alterations in the shape of the vocal tract as articulators are manipulated during the production of various speech sounds. Pickett (1999) defined formant frequencies as resonances in sound transmission through the vocal tract. The formants of the speech syllable /u/ are represented in the spectrogram below. The dark bands represent formant frequencies (i.e., concentration of acoustic energy on top of harmonics).

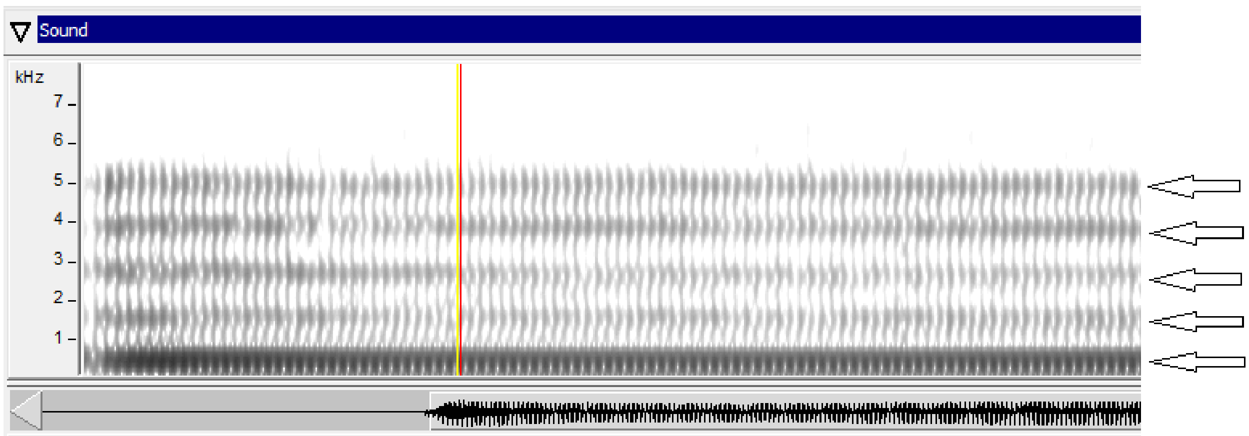

Figure 4. Wideband spectrogram with better temporal resolution showing formants (indicated by arrows).

A thorough understanding of these three speech acoustic components is essential to accurately record and interpret the FFR. Each of these acoustic components is encoded by the human auditory brainstem, and these components will be the subject of most of the FFR analysis and interpretation discussed in this project.

Auditory Neural Encoding

The auditory nerve and auditory portion of the brainstem have distinct neurophysiological properties that are unique and that do not exists in other neurological areas in the body. The next section will explain two of these neurophysiological properties.

Phase-locking. Phase-locking refers to the preferential firing of neurons at a discrete phase of an amplitude-modulated stimulus (Koppl, 1997). In other words, when an auditory nerve is “phase-locked”, it only fires at certain points in time, which are related to the period of the stimulus. For example, if a 500 Hz tone is presented, the auditory nerve fires at whole number multiples of the period of the stimulus:

P=1000 ms/500 Hz

P=2 ms

The auditory nerve fires at 2 ms, 4 ms, 6 ms, 8 ms, etc…

Envelope & temporal fine structure. The human auditory system is able to process both basic sounds (pure tones) and complex sounds (speech). Multiple complex processes underlie auditory perception of complex speech sounds. This section will go into brief detail about each of these processes.

When a signal reaches the inner ear, the cochlea begins analyzing and disintegrating the entire signal (broadband) into parts via auditory filters along the basilar membrane. Lorenzi, Gilbert, Carn, Garnier, and Moore (2006) have described the output of these auditory filters as a “bandpass filtered version” of the input signal. Moore (2008) explained that when complex signals are processed by the auditory system, there is a breakdown of the signal into two components, the slowly-varying E and the TFS. The component with slow variations in amplitude (magnitude) over time is known as the E. This envelope of the complex waveform is modulated at the F0. On the other hand, the component with rapid variations/oscillations that has a rate similar to that of the center frequency of the input signal is known as the TFS. The underlying fine structure component represents the higher frequency content of the signal. In the present study, we will refer to the F0 when we use the term envelope. The TFS will be used to describe the higher formants. Both the E and TFS encoding ability that is observed at the level of the auditory nerve is preserved at the level of the brainstem as evidenced in the neural activity of the brainstem FFR. In the next section of the literature review I will discuss the history of this electrophysiologic response.

History of the FFR

The history of the FFR is relatively short, with roughly only 40-50 years of research in the area. Worden and Marsh initially described the FFR in 1968. For the first time, an evident distinction in the AEP literature was made between the cochlear microphonic (CM) measured using an electrocochleography technique and the FFR. Prior to 1968, the CM was believed to be the only auditory response capable of reproducing/mimicking the temporal waveform and frequency properties of the stimulus used to elicit the response. Artifact, such as acoustic vibration, was also at one point believed to be responsible for this reproduction (Marsh, Worden, & Smith, 1970). Therefore, to truly understand the background of the FFR, the history of the CM must first be understood.

Wever and Bray (1930) proposed that when an electrode was placed on the auditory nerve (proximal to the medulla) of a cat and sound stimuli (either a pure tone or speech) was applied to an ear, a reproduction of "great fidelity" could be recorded. That is, either pure-tones or speech could be presented, recorded, played back, and identified. This effect was found to be audible for stimulus frequencies up to 3300 Hz. These investigators wanted to ensure that an artifact of the stimulus was not the potential generator of this CM response. In order to do this they placed the active electrode on other central nervous tissue (medulla, auditory brainstem), and reported that when this was done, no response was obtained suggesting that the electrophysiologic response was most likely not due to stimulus artifact. Wever and Bray also reported that when the cochlea was destroyed on the side where the electrode was placed, the response amplitude of the CM was significantly decreased. When the opposite cochlea was also destroyed, the CM was completely erased. This evidence opened the door to the possibility of a response that follows the acoustic properties of the stimulus being generated at levels beyond the structure of the cochlea.

Researchers Worden and Marsh (1968) expanded on the initial findings of Wever and Bray (1930), and are credited with coining the term the "frequency following response" since they observed microphonic-like responses coming from the auditory nerve in response to sounds. This study was seminal in identifying and separating the FFR from previous suggestions that related it to the CM. Worden and Marsh studied both the onset of the FFR and the latency of this response and compared it to the properties of the CM. First, they reported that the onset of the CM was "graded" or gradual, whereas the onset of the FFR was sudden and abrupt. Secondly, the latency of the onset of the FFR was delayed by 5-10 ms, revealing that the generator site or origin of the response was likely well beyond the cochlea.

Two years later, Marsh et al. (1970) performed two additional experiments to rule out the cochlea as a generator site of the FFR. These investigators were attempting to further distinguish the FFR from the CM. In the first experiment, the auditory nerve was intentionally severed at the level of the cochlear nucleus (CN) to study any changes in the

FFR. When the auditory nerve was severed in animal subjects, the FFR was completely destroyed, but the CM was recordable. This again is a strong piece of evidence suggesting that the FFR in humans is generated by structures in the CANS beyond the cochlea. Figure 5 below illustrates the FFR and CM before (sections A + B) and after (sections C+D) the sectioning of the auditory nerve.

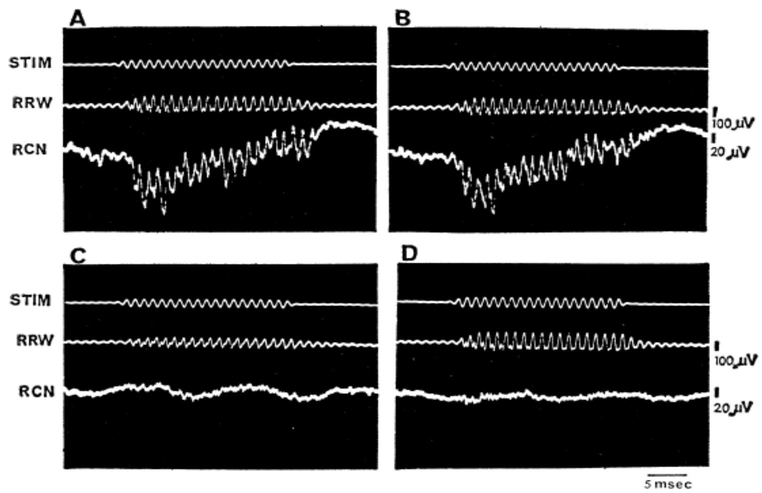

Fig. 1. Effects of transection of the eighth nerve on the cochlear microphonic (CM) and the frequency-following response (FFR). In each photo traces are: *STIM*, electrical signal to earphone produces a tone pulse with a frequency of 730 hz, duration of 25 msec, and an intensity of 100 db referred to .0002 µbar; *RRW*, cochlear microphonic recorded at the right round window; *RCN*, response recorded from right cochlear nucleus. The photos in the top row are (A) taken 5 minutes before, and (B) just before section of the eighth nerve. Photos in the lower row are (C) taken immediately after, and (D) 5 minutes after section of the eighth nerve. The CM response survives the section, whereas the FFR and the evoked potential from the cochlear nucleus are abolished.

Figure 5. The FFR (RCN) and CM (RRW) before and after being sectioned. Adapted from "Auditory Frequency-Following Response: Neural or Artifact?" J.T. Marsh, F.G. Worden, and J.C. Smith, 1970, *Science,* 169(3951), 1222-1223.

In the second experiment, a "reversible blocking" technique which involved cooling of the CN was used to study variations in the FFR. Both the FFR and the cochlear microphonic were recorded at three various times (i.e., before the CN was cooled, at a cooled state, and after being cooled). Marsh et al. (1970) reported that the FFR was abolished when the CN was cooled but recovered when the temperature returned to its normal state (as seen in the Figure 6 below). In contrast, the CM showed no changes in response to the cooling. This is yet another strong piece of evidence that suggested that the CM and the FFR have separate underlying generators.

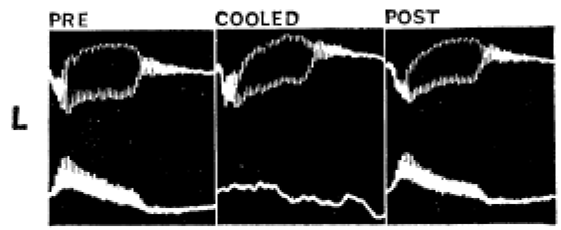

Figure 6. Cochlear nerve response to “reversible blocking” technique. From left to right: cochlear nerve activity prior to technique; cochlear nerve activity during technique application; cochlear nerve activity following technique. Adapted from “Auditory Frequency-Following Response: Neural or Artifact?” J.T. Marsh, F.G. Worden, and J.C. Smith, 1970, *Science*, 169(3951), 1222-1223.

The FFR was first described in human subjects by Moushegian, Rupert, and Stillman (1973), who recorded this response in five normal hearing participants. Moushegian et al. (1973) recorded the FFR to several different stimulus frequency tones (i.e., 250 Hz, 500 Hz, 1000 Hz, 1500 Hz, and 2000 Hz). The FFR was recorded in both a quiet as well as a masked condition. Similar to previous findings of the FFR in animal subjects, Moushegian and colleagues reported that the peaks in the FFR occurred at intervals equal to the period of the stimulus. This was true for each stimulus frequency assessed. A decrease in response amplitude was present at 2000 Hz, but the response was still found to follow the period of the stimulus.

In the Moushegian et al. (1973) experiment, the FFR was also recorded at various stimulus intensities at each stimulus frequency. These stimulus intensities were referenced to the patient’s behavioral thresholds at the specific frequencies. The sensation levels were 5 dB, 11 dB, 26 dB, 36 dB, 46 dB, and 56 dB. Moushegian and colleagues reported that some type of response could be identified at sensation levels between 10-20 dB; however, the FFR specifically required a sensation level of at least 46 dB. These investigators also reported that the latency between the onset of the stimulus and the FFR was found to occur at approximately 6.0 msec. At a latency of 6.0 msec, the response was suggested as being of neural origin and therefore not the result of acoustical artifact. The onset latency of the FFR was similar to that seen in previous animal studies (Marsh et al., 1970; Wever & Bray, 1930; Worden & Marsh, 1968). Lastly, when continuous noise masking was applied to mask out contributions from the cochlea, the FFR could not be recorded. In the quiet (no noise masking) test condition, the FFR was present. Collectively, this evidence appears to confirm that the FFR is neural in origin.

Based on this collective evidence from human and animal studies, the FFR is widely accepted as a neural response. Focus of research in the area of the FFR has shifted from understanding its origin and characteristics to investigating its clinical applicability. Specifically, the FFR has been proposed for several purposes including: evaluating the integrity of physiological mechanisms responsible for the encoding and processing of complex sounds such as speech; studying the effects of cochlear impairment on subcortical representation of speech; such as the link between subcortical representation and perception of speech; and the effects of neural plasticity on the FFR (Burkard, Eggermont, & Don, 2007). The next section will focus on explaining the FFR in more detail.

What is the FFR?

The Frequency Following Response (FFR) is an AEP that responds to stimuli which exhibit repetitive periodic features (Skoe & Kraus, 2010). This response is phase-locked to the stimulus which elicits it. As a result, there is a “mimicking” or “following” of the stimulus frequencies. In the FFR, the distance between the peaks of the response is equivalent to the period of the fundamental frequency of the stimulus. A basic representation of an FFR to a 500 Hz tone is detailed in the illustration below.

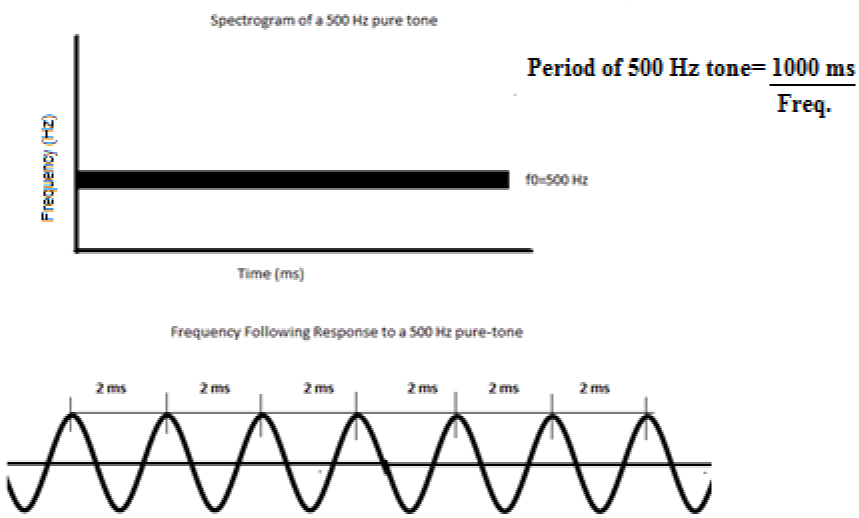

Figure 7. The 500 Hz tone has a F0 equal to 500 Hz (represented in the spectrogram). The peaks in the temporal waveform of the FFR occur every 2 ms, which corresponds to the period of the fundamental frequency or F0.

The break-down of any complex stimulus into envelope and TFS occurring at the auditory nerve level is preserved in the neuronal activity in the rostral brainstem underlying the FFR. Therefore, the FFR obtained in response to any complex stimulus reflects phase locking to both envelope and TFS information. The envelope refers to the sub-cortical representation of the F0 of the stimulus. Speech signals, especially vowels, are characterized by a complex frequency spectrum consisting of both a F0 as well as a rich harmonic structure. The subcortical representation of the harmonics is known as the fine structure component of the FFR.

A brief discussion of stimulus polarity (further described in the stimulus parameters section) is needed before addressing analysis of the FFR to E and TFS. Polarity refers to the initial displacement of the stimulus. Rarefaction is defined as an initial negative polarity of the stimulus; whereas condensation is defined as an initial positive polarity of the stimulus. An alternating polarity stimulus employs both condensation and rarefaction polarities. When the FFR is recorded to a stimulus presented in a single *polarity*, such as condensation or rarefaction, the resultant temporal waveform contains both the envelope and fine structure components. In order to tease apart the neural phase-locking to envelope from the neural phase-locking to the TFS, the FFR should be recorded using an alternating polarity stimulus as a first step, which then can be separated into the response to the rarefaction stimulus and the response to the condensation stimulus (Krishnan, 2007; Russo, Nicol, Musacchia & Kraus, 2004; Wile & Balaban, 2007). When the rarefaction recording is added to the condensation recording, the envelope temporal response is produced. Conversely, when the rarefaction recording is subtracted from the condensation recording, the brainstem neural response to the fine structure remains.

In order to analyze neural encoding to each frequency component of the complex FFR, the temporal waveform must be converted into the frequency domain. This is achieved by using a Fast Fourier Transform (FFT) analysis, which mathematically converts a response from the temporal domain into the frequency domain (Katz, 2009). Both the E and TFS of the FFR display characteristic differences in the frequency domain. In the FFT of the TFS component, there are multiple peaks of the energy which occur at the harmonic frequencies of the F0 (as seen in Figure 9).

In contrast, in the E component, the FFT reveals a single peak, which indicates that energy is present only at the F0 (as seen in Figure 10). The FFT of the responses essentially shows precise replication of the stimulus by the auditory system at the level of the brainstem. Below is a very basic example of what both FFT's would look like for the FFR elicited by a tone complex. For theoretical purposes, let's imagine that the fundamental frequency of the speech stimulus is 250 Hz and the harmonics are 500 Hz (F2), 750(F3), 1000 Hz (F4). This is illustrated below:

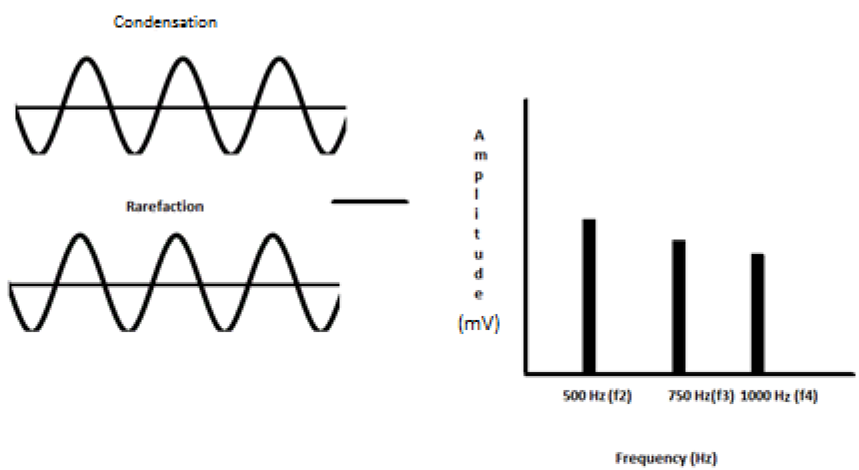

Figure 8. Subtracted waveform is obtained when the FFR obtained to a condensation stimulus polarity is subtracted from the FFR obtained to a rarefaction stimulus polarity. The FFT of the subtracted waveform is illustrated on the right side. This FFT shows energy which corresponds to the temporal fine structure (whole number multiples of F0).

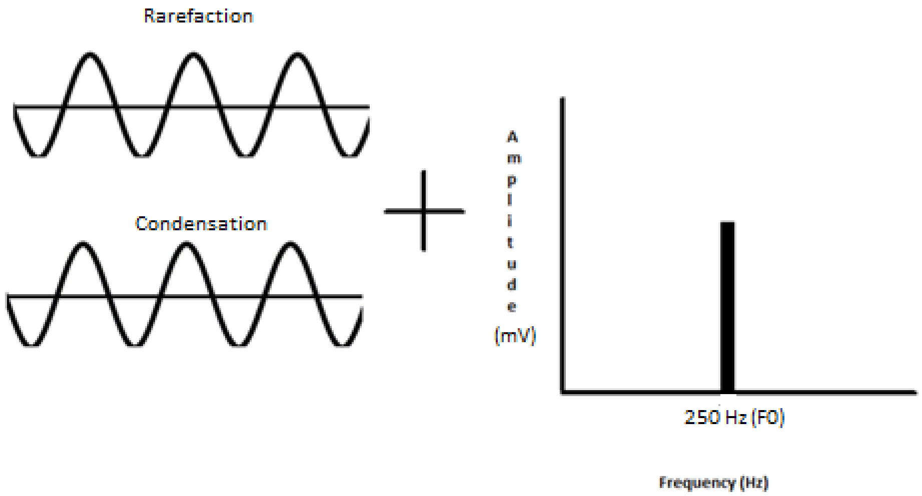

Figure 9. Summed waveform is obtained when the FFR obtained to a rarefaction stimulus polarity is added to the FFR obtained to a condensation stimulus polarity. The FFT of the summed waveform is illustrated on the right side. This FFT shows energy which corresponds to the envelope (F0).

Aside from the importance of understanding what the FFR is, recognizing the neural origins of this response is just as essential for selecting appropriate recording parameters. The next section will go into detail about various studies that investigated the neural generators of the FFR.

Neural Generators of the FFR

The exact underlying neural generators of the FFR are debatable; however, various locations at the level of the auditory brainstem have been proposed (Gardi, Merzenich & McKean, 1979; Smith, Marsh, & Brown, 1975). Smith et al. (1975) conducted two different experiments to determine what structures within the auditory brainstem pathway were the sources of the FFR. In the first experiment, the FFR was recorded in both a near-field test condition (with needle electrode located on the cochlear nucleus (CN), superior olivary complex (SOC), and inferior colliculus (IC) in cats) and in a far-field recording testing condition (electrodes on various locations on scalp in humans). The onset latency of the FFR was then compared between these two test conditions. The table below displays the mean FFR onset latencies in the near-field recording (top portion of table) compared to far-field test condition (lower portion of table).

TABLE 1

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

Figure 10. FFR onset latencies of near-field recordings at various acoustic brainstem sites in cats and scalp far-field recording in humans. Adapted from "Far-field recorded frequency-following responses: evidence for the locus of brainstem sources," by J.C. Smith, J.T. Marsh, and W.S. Brown, 1975, Electroencephalography and Clinical Neurophysiology, 39(5), 465-472.

As can be seen, the mean latency recorded at the level of the inferior colliculi in seven cats (5.4 msec) was most approximate to the mean far-field FFR latency recorded at the scalp in both cats and humans (5.8 ms in cats and 6.5 ms humans). This finding suggests that the inferior colliculus is a likely neural generator of the FFR for both cats and humans.

The second experiment in the Smith et al. (1975) study focused more specifically on the IC. A chemical agent known as "reversible cryogenic blockade" was used to cool and/or warm the IC and to determine any effect on the amplitude of the far-field recorded FFR. It has been demonstrated that this chemical agent does not have any effect on any of the lower brainstem nuclei. When cooling of the IC was performed, the vertex recordings were either diminished in amplitude or eliminated. In contrast, when the IC was warmed the amplitude of the FFR reverted back to normal. Interestingly, the medial olivary complex (MOC) showed no changes to the warming or cooling effect of the cryogenic blockade.

Smith and colleagues concluded that the primary source of the FFR was the IC and not multiple brainstem sources. This conclusion was based on two pieces of evidence: (1) the near-field recording at the IC approximating closest to the onset latency of the FFR in far-field scalp recordings, and (2) only the recordings at the IC showing an effect to cooling and warming to cryogenic blockage chemical agent.

In 1979, Gardi, Merzenich, and McKean investigated the origin of scalp recorded FFR in cats. In contrast to the findings from Smith et al. (1975), these investigators reported that the IC played a much less significant role in generation of the FFR. Instead, Gardi et al. (1979) reported that the FFR is the result of contributions from more than a single neural generator. Specifically, they reported that the primary neural generator of the response was the CN, which contributed to 50% of the amplitude of the FFR. The other auditory structures involved were the SOC (20%) and the cochlea (25%). Gardi et al. (1979) proposed that because there are multiple generators that contribute to the FFR, the amplitude of this response varies when the stimulus frequency is changed.

At a similar time period, Stillman, Crow, and Moushegian (1978) investigated the components of FFR in human participants and also found there to be more than one neural generator. In this study, FFRs were recorded from the scalp using two different electrode montages, a horizontal montage and a vertical montage. A horizontal montage is when electrodes are located on both earlobes/mastoids. In a vertical montage, the reference electrode is located on the earlobe and the non-inverting is located at the vertex or Cz. Two electrode montages were used in this study because a horizontal montage captures energy from peripheral structures such as the auditory nerve, whereas a vertical montage captures energy from more central auditory structures. Therefore when the responses from these two montages were directly compared, the investigators were able to study the contributions from different peripheral and central auditory structures to the response.

Based on the recordings from the two electrode montages, Stillman et al. (1978) reported the existence of two different FFRs, which they labeled FFP1 and FFP2. The FFP1 response could be acquired using both electrode montages (horizontal and vertical). In contrast, the FFP2 could only be recorded with the vertical electrode montage. Also the FFP2 could be elicited at a lower stimulus intensity in comparison to the FFP1, (i.e., 10 dB below the stimulus intensity needed to elicit FFP1 using the horizontal electrode montage.) Collectively, this evidence suggests that the primary neural origin of the FFP2 response was from a more central auditory brainstem structure (IC) where the FFP1 has a more peripheral origin (CN/SOC). As a result of these findings, Stillman et al. (1978) proposed that two different regions (IC vs. CN/SOC) within the auditory brainstem are capable of producing the FFR.

More recently, Galbraith (1994) furthered the notion proposed by Stillman et al. (1978) of the existence of two different responses within the FFR. Using both a horizontal and a vertical electrode montage, these investigators recorded the FFR to a 200

Hz pure tone as well as to a complex stimulus with a missing fundamental frequency. Based on the latency characteristics of the FFR recorded using both electrode montages and an analysis of both the envelope and fine structure of the FFR, Galbraith suggested that there are two areas within the brainstem that contribute to the response. These two areas are: the rostral portion of the brainstem, specifically the lateral lemniscus and IC, and the caudal portion of the brainstem, specifically the CN. These findings are in good agreement with those of Stillman et al. (1978).

In 2010, Chandrasekaran and Kraus have also suggested that there is more than one neural generator of the FFR. Both the cochlea and auditory cortex have been ruled out as contributors to this response. Today, the cochlear nucleus, lateral lemniscus, and inferior colliculus are all nuclei within the auditory brainstem which are widely accepted as neural generators of the FFR (Chandrasekaran & Kraus, 2010). Now that we understand the neural origins of the FFR, the next section will discuss appropriate technical parameters needed to capture the response from these structures.

FFR Technical Parameters

The FFR as well as other AEPs can also be affected by a number of technical parameters such as stimulus, recording, and subject parameters. Stimulus parameters describe the characteristics of the stimulus that include and are not limited to intensity, rate, polarity and stimulus type. Acquisition parameters describe the settings/protocols of the recording equipment that are unique to each AEP and which must be correctly set to accurately capture the response of interest. These settings/protocols include electrode montage, sampling rate, length of averaging window, number of trials, band-pass analog filter settings, and artifact. Subject parameters are concerned with subject-related factors which can alter or impact the ability to successfully record AEPs. In the case of the FFR, two known subject-related factors are age and attention/state. The following section will address the various technical parameters which are specific to the FFR.

FFR Stimulus Parameters

There are many stimulus parameters that can affect the ability to successfully record the FFR. These stimulus parameters are intensity, rate, polarity, and type of stimulus. Each of these stimulus parameters will be discussed in the following section.

Stimulus type. The FFR can be recorded to a variety of acoustic stimuli. These acoustic stimuli can be classified in a number of ways (as illustrated in figure 12 below).

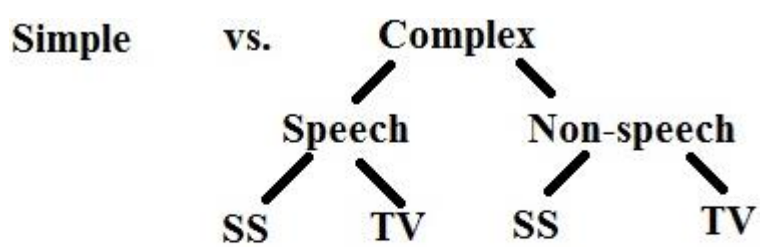

Figure 11. Classification system of simple and complex acoustic stimuli (SS= Steady-state, TV= Time-varying).

The first way of classifying acoustic stimuli refers to their spectral waveform. That is, whether they are simple (pure tones) or complex sounds (for distinction please see basic acoustics section). Complex stimuli can be subsequently broken down into two types of sounds: speech and non-speech stimuli. Both speech and non-speech stimuli can be further defined by their spectral characteristics (i.e., harmonics, formant frequencies, fundamental frequency). A complex stimulus that has a stable F0 and harmonic structure over the course of its total duration is known as a steady-state stimulus (SS) (as seen on left side of figure 13). On the other hand, a complex stimulus that does not have a stable F0 and harmonic structure over the course of its total duration is known as time-varying stimulus (TV) (as seen on right side of figure 13).

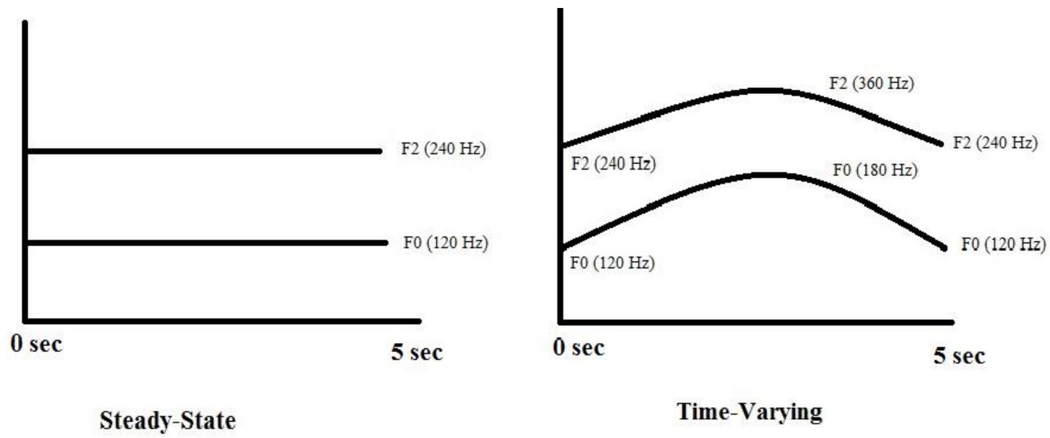

Figure 12. Steady-state and time-varying representations of complex stimuli with a duration of five seconds.

In the steady-state stimulus, as the name implies, the F0 (120 Hz) and second harmonic (240 Hz) remain the same for the entire duration of stimulus (5 sec). In contrast, in the time-varying stimulus the F0 and second harmonic change over the duration of stimulus. Whenever the F0 changes, so do the harmonics (formants also change as F0 changes). The changes in the harmonics are proportional to the F0.

As previously mentioned, speech stimuli can be either steady-state or time-varying or both. For example, the Chinese syllable /ma/ can be uttered using a total of four pitch contours (mā, má, mǎ, mà). As Chinese is a tonal language, the meaning of the syllable /ma/ changes with each pitch contour. When /ma/ is uttered with no rising or falling intonation, it means mom (mā). When it is uttered with a falling intonation, it means to scold (mà). The Chinese speech stimulus /mā/ represents a steady-state stimulus whereas the stimulus / mà / is an example of a time-varying stimulus, with (rising/falling)

F0 and harmonics. More examples of time-varying speech stimuli include diphthongs, consonant-vowel formant trajectories, musical glissandos and other tonal language pitch contours (e.g., Vietnamese). An example of complex non-speech steady-state stimuli are musical notes (e.g., chords). Time-varying non-speech stimuli include background noise, environmental sounds, and amplitude modulated (AM) tones.

The types of speech sounds employed in FFR studies have included CV syllables, consonants, vowels, and Chinese language tones. The types of non-speech sounds used for FFR recordings have included musical notes, background noise, and amplitude-modulated tones. In 1995, Levi, Folsom and Dobie utilized amplitude modulated (AM) tones with carrier frequencies of 500 Hz, 1000 Hz, and 2000 Hz to study the FFR in normal hearing infants and adults. Six years later, Cunningham, Nicol, and Zecker were interested on the effects of noise on the FFR and employed a synthetic /ada/ speech stimulus with "conversational" or background noise characteristics. More recently, there has been an interest in the selection of musical notes for studying the FFR. Bidelman and Krishnan (2010) used iterated rippled noise (IRN) with musical characteristics (major third). Around the same time, Bidelman and Krishnan (2010) developed and used a synthetic /i/ speech stimulus with varying degrees of reverberation in an attempt to study the effect of reverberation on the FFR and compare differences between individuals with and without musical experience. The CV syllable /da/ has also been extensively used in various FFR studies to investigate auditory brainstem plasticity, speech encoding in children with learning disabilities, musical experience in the presence of background noise and more (Cunningham et al., 2001; King, Warrier, Hayes, & Kraus, 2002;

Musacchia, Sams, Skoe, & Kraus, 2007; Parbery-Clark, Skoe, & Kraus, 2009; Russo et al., 2005).

As the major focus of the present study is on subcortical encoding of speech in humans in adverse listening conditions, and given the success of speech stimuli in eliciting robust FFRs, the stimulus in the current experiment will be the speech stimulus /u/.

Stimulus intensity. Some of the earliest FFR studies looked at the relationship between stimulus intensity, the FFR, and behavioral thresholds (Davis & Hirsh, 1976; Moushegian et al., 1973). In 1973, Moushegian and colleagues recorded the FFR to several different stimulus frequency tones that ranged from 250 Hz-2000 Hz. The FFR to each stimulus frequency tone was recorded at six different intensity levels that referenced participants' behavioral thresholds (e.g., 5 dB SL, 11 dB SL, 26 dB SL, 36 dB SL, 46 dB SL, 56 dB SL). Researchers in this study found that while the FFR could be somewhat identified at stimulus intensities within 10-20 dB of participants' behavioral thresholds, a minimum stimulus intensity of 46 dB was required in order to obtain a temporally precise FFR.

Three years later, Davis and Hirsh (1976) recorded the FFR in response to a low frequency tone burst (500 Hz) in 21 normal hearing subjects. The FFR was recorded to five different stimulus intensity levels referencing participants' behavioral thresholds in 10 dB steps (e.g., 30 dB-70 dB). These researchers reported that the 40 dB SL represented the FFR "threshold", and an "on-effect" could be identified at this level. However, they suggested that at higher stimulus intensities (60-70 dB SL), the FFR was smoother and more temporally precise than at stimulus intensities of 50 dB SL and below.

More recently, Krishnan (2002) investigated the effect of stimulus intensity on the FFR recorded to steady-state synthetic vowels in 10 normal hearing participants. Stimulus intensities were varied in 10 dB steps from 55 dB – 85 dB nHL. An intensity effect was observed as the stimulus presentation level was varied from 55 to 85 dB nHL; specifically as stimulus intensity was increased there was an increase/improvement in both temporal and spectral-domain representations of the subcortical neural signal, as pictured in Figure 14 (Krishnan, 2002). More specifically, both the amplitude of the time waveform of the FFR, as well as the magnitudes of the components at the F0 and first formant increased with increasing stimulus intensity.

A similar study was carried out by Akhoun, Gallego, and Moulin (2008) in which the FFR was recorded to a speech stimulus (/ba/) at intensities ranging from 0-60 dB SL (presented in 10 dB increments) in a group of normal hearing listeners. When the stimulus intensity was between 0-50 dB SL, the pattern of the FFR was relatively flat (decreased waveform amplitude) compared to the pattern of the FFR when a stimulus intensity of 60 dB SL was used. As seen in Figure 15 (Akhoun et al., 2008), the FFR recorded at 60 dB SL most closely represented the envelope of the speech stimulus /ba/. As stimulus intensity was increased the latency of both the onset response and the FFR decreased.

Collectively, the results of these studies clearly demonstrate that subcortical neural responses exhibit greater response amplitudes, temporal precision, and spectral clarity at supra-threshold levels of about 60-85 dB SPL compared to lower stimulus presentation levels. Therefore in the present study we will employ a stimulus intensity level within this range.

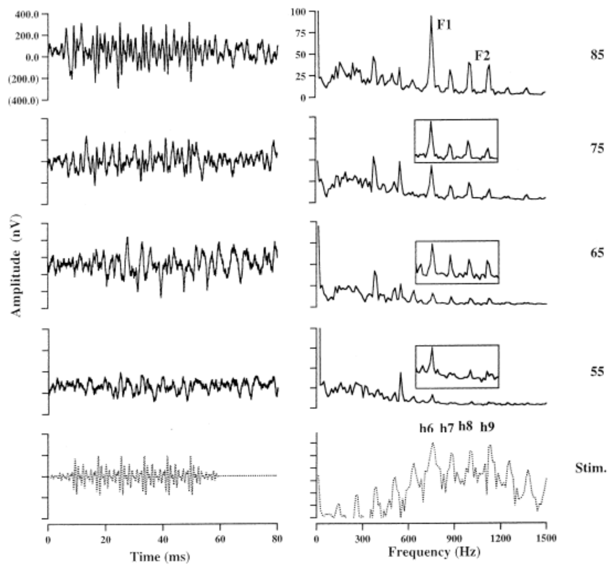

Fig. 5. Grand averaged FFR waveforms and spectra are plotted as a function of stimulus intensity (dB nHL) for the vowel /a/. The stimulus waveform and its spectrum (with F_1 and F_2 harmonics identified) are at the bottom of each panel. The amplified inset in the FFR spectral data clearly shows the F_2 harmonic peaks. Note the different amplitude scale for the stimulus spectrum.

Figure 13. FFR waveform and spectra amplitude at various stimulus intensity levels. Adapted from "Human frequency-following responses: Representation of steady-state synthetic vowels," by A. Krishnan, 2002, *Hearing Research*, 166:192–201.

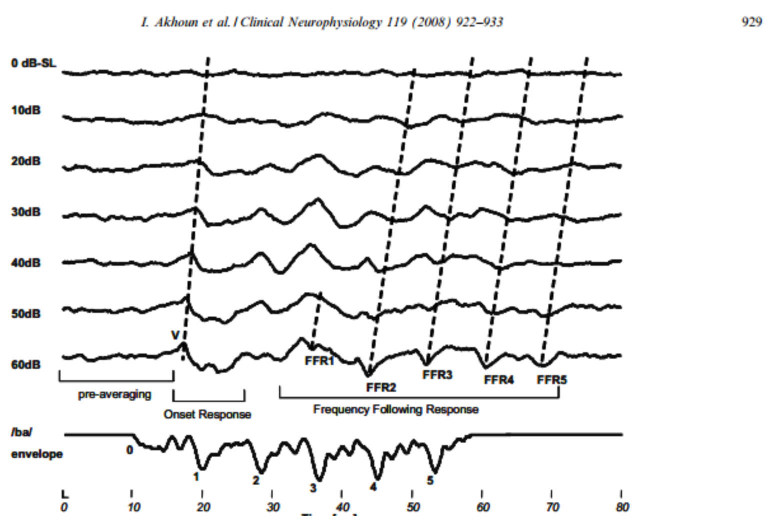

Fig. 5. Speech ABR latency/intensity function. The wave V is supposed to be elicited by the beginning of the stimulus and each FFR wave is supposed to be elicited by the corresponding peak of the envelope (FFR1 generated by '1', and *idem* for FFR2 to FFR5). The bottom waveform is the envelope of the /ba/ (*idem* Fig. 4). Note that the latency of each component (wave V, FFR 1, FFR 2, FFR 3, FFR 4) clearly increases as intensity decreases. The grand-averaged Speech ABR shown here was obtained for alternate polarity in 11 subjects.

Figure 14. FFR recorded to speech stimulus /ba/ at various sensation levels. As intensity level increased, the latency of FFR decreased. Adapted from "The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults," by I. Akhoun, S. Gallego, and A. Moulin, 2008a, *Clinical Neurophysiology*, 119:922–933.

Stimulus Rate

In the FFR literature, stimulus rate is discussed in the context of either interstimulus interval (ISI) or stimulus onset asynchrony (SOA). Interstimulus Interval (ISI) refers to the period of silence, in milliseconds, between the offset of the preceding stimulus and the onset of the following stimulus (as seen on left side of figure 16). In contrast, Stimulus Onset Asynchrony (SOA) reflects the total duration of the stimulus as well as the period of silence that follows the stimulus and precedes the onset of the following stimulus (as seen on right side of figure 16). ISI and SOA of a tone burst stimulus are illustrated in figure 16 below.

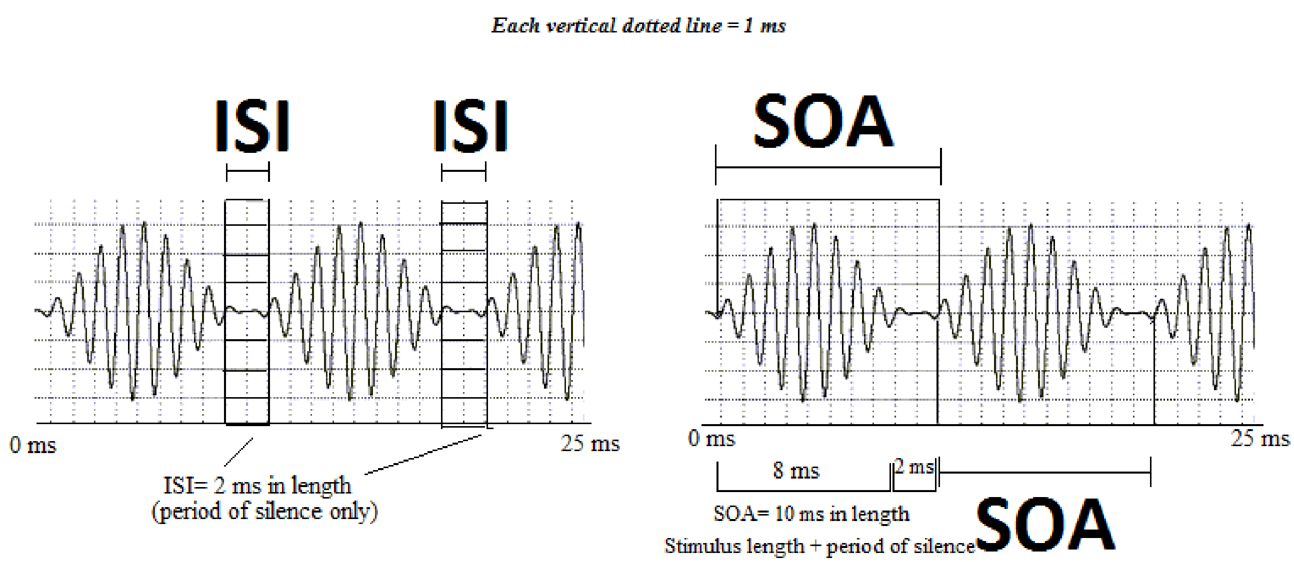

Figure 15. Schematic comparison of length of duration between ISI and SOA for stimulus rate.

In this example, the ISI is 2 ms in length as it only represents the period of silence between the two stimuli. The SOA is 10 ms in length as it includes both the total stimulus duration (8 ms) and the period of silence following the stimulus (2 ms).

According to Skoe and Kraus (2010), the ISI used in multiple FFR studies has ranged from 30% of the stimulus length to over double the stimulus length. Three main factors must be taken into account when determining the ISI (Skoe & Kraus, 2010). First, one must understand the impact of an ISI on the auditory perception of a sound, especially complex sounds such as speech. According to Hall (2007), the impact of the ISI on auditory perception is directly related to “basic neuro-physiologic mechanisms.” These mechanisms refer to the period following nerve excitation known as the recovery or absolute refractory period. During this time, a neuron is unable to fire and its excitation threshold is much higher. The ISI is directly related to this absolute refractory period. For example, if the ISI is longer than the refractory period, than the auditory neuron can fully recover and a robust response can be acquired. However, if an ISI is shorter than the refractory period, then the response is less robust and there usually is an increase in latency and a decrease in response amplitude.

Next, an ample pre-stimulus averaging window must be selected; since one must be certain that the preceding stimulus can fully conclude before the following stimulus is presented. Having a pre-stimulus averaging window of appropriate duration ensures the return of the response from the peak to the baseline, thus making the analysis of signal-to-noise (SNRs) easier and preventing neural adaptation. Neural adaptation is decreased responsiveness of a sensory system in response to a constant or repetitive stimulus. Because the FFR is recorded to repetitive stimuli, it is essential that the selected ISI is large enough to prevent modulation of the response.

Finally, one must consider the actual stimulus rate. Stimulus rate refers to how many times a stimulus is presented over a discrete time period, typically one second.

Thus, a stimulus that is presented 3 times over the course of one second has a stimulus rate of 3/sec. In ensuring that there is no contamination of the stimulus from an AC line frequency (60 Hz), a non-integer presentation rate was recommended by Skoe and Kraus (2010)(i.e., using a rate of 10.3.sec instead of using a rate of 10/sec or 20/sec, which could potentially confound the response). The stimulus rates employed in recent studies using harmonic complexes or speech stimuli are listed in the Table 1 below. All of these studies have employed rates well over double the stimulus length. The stimulus rate can be calculated using the equation (1000 ms/ISI), since ISI is the reciprocal of stimulus rate. In this case, 1000 ms divided by an ISI of 319 yields a rate of 3.13/sec. Keeping these factors in mind, the present study will employ a stimulus presentation rate of 3.13/sec, which corresponds to an ISI of 319 msec.

Table 1

Stimulus Duration and Rate of Multiple FFR Studies

	Krishnan, Xu, and Gandour (2004)	Gockel, Carlyon, Mehta, and Plack (2011)	Gnanateja, Ranjan, and Sandeep (2012)	Jeng et al. (2011)
Type of Stimuli	Speech stimulus /yi/	Harmonic complex with an F0 of 244 Hz	Speech stimulus /da/	Speech Stimuli /i/ (117-166 Hz)
Stimulus Duration	250 ms /0.25 sec	100 ms/ 0.10 sec	40 ms/ 0.04 sec	250 ms/ 0.25 sec
Stimulus Rate	**3.13/s**	**3.57/s**	**10.9/s**	**3.39/s**

Stimulus polarity. As mentioned and explained on page 19, the FFR can be recorded using condensation, rarefaction, or alternating polarity stimuli. Manipulating the stimulus polarity allows the clinician to separate the E and the TFS components of the FFR to quantitatively and qualitatively analyze subcortical encoding of the F0 and formant structure. The manipulation of stimulus polarity for the purpose of enhancing/minimizing spectral components or envelope has its origin in the compound histogram technique, which was first proposed by Goblick and Pfeiffer (1969) and further developed by Arthur, Pfeiffer, and Suga (1971).

Researchers Goblick and Pfeiffer (1969) first identified the use of two stimulus polarities (rarefaction + condensation) as a 'nulling technique' since two different responses were found to exist when cochlear nerve fibers responses were recorded and polarities were either added or subtracted. Two years later, Arthur et al. (1971) investigated the properties of two-tone inhibition in primary auditory neurons in response to phase-locked sound stimuli (tone-bursts) and utilized this method to analyze 'discharge patterns' of the response. With regards to recording the FFR to both stimulus polarities for separating and analyzing E and TFS components, Krishnan (2002) was the first to do so in an investigation of how steady-state vowels are represented in the human FFR. This study was specifically interested in the spectral components of the FFR. In order to minimize the E component of the FFR and optimize the spectral components, the FFR recorded to each stimulus polarity was subtracted from one another. Following this subtraction, an FFT was performed on each of the difference in responses to visualize the desired spectral peaks.

More recently, Aiken and Picton (2008) have explained that differentiating between FFR components can be difficult unless the FFR is recorded using an alternating polarity. In using a stimulus with an alternating polarity, both components can be separated and studied by adding or subtracting the responses to the same stimulus recorded in opposing polarities. In doing so, response components of the envelope of the

FFR can be separated from the response components of the spectrum of the FFR (Aiken & Picton, 2008). The addition and subtraction of FFR responses for purposes of separating and enhancing specific components has been successfully performed by many previous studies (Akhoun et al., 2008; Krishnan, 2002; Russo et al., 2004; Wile & Balaban, 2007). Table 2 below explains the average response nomenclature for separating FFR components recorded to an alternating polarity stimulus (from Aiken & Picton, 2008). Given our interest in analyzing both the spectral FFR (fine structure) and the envelope FFR, the current investigators will use an alternating polarity stimulus.

Table 2

Average Response Nomenclature

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

Note. Adapted from “Envelope and Spectral Frequency-following Responses to Vowel Sounds,” by S.J. Aiken and T.W. Picton, 2008, *Hearing Research, 245 (1),* p. 35-47.

Recording Parameters

There are a number of recording parameters that need to be understood in order to successfully obtain a reliable FFR. These recording parameters include electrode montage, sampling rate, length of averaging window, number of trials, band pass analog filter settings, and artifact. Each of these recording parameters is discussed in the section below.

Electrode montage. The FFR can be recorded using either a vertical or a horizontal electrode montage. In a three-channel vertical montage, an active electrode is located on Fz, a reference/inverting electrode is located on the mastoid (M1, M2), and on the seventh cervical vertebrae at the back of the neck (C7), and a ground electrode is located on the nasion (Fpz). A vertical two-channel montage might consist of linked mastoids (M1, M2) serving jointly as one reference electrode with the seventh cervical vertebrae (C7) acting as the second reference On the other hand, a vertical three-channel montage may consist of one active and three reference electrodes, on M1, M2, and C7 respectively. The figure below illustrates a vertical electrode montage.

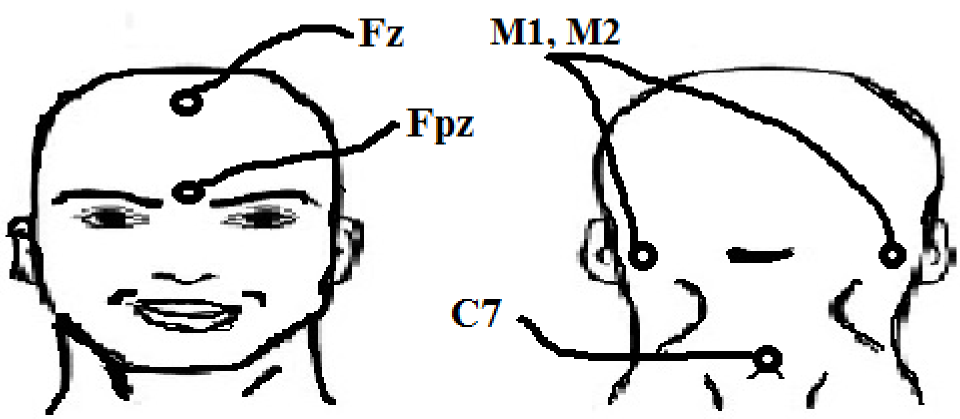

Figure 16. Example of a typical three-channel vertical montage used to record the FFR.

Galbraith (1994) described the advantage of employing a vertical montage over a horizontal montage. He reported that for recording the FFR, the horizontal montage captured the energy mostly from peripheral structures, especially contributions from the auditory nerve. Whereas, the vertical montage captured contributions to the response from more central structures within the auditory brainstem. It has been demonstrated that several of these higher brainstem structures such as the lateral lemniscus (LL) and IC, are known generators of the FFR (Galbraith, 1994; Smith et al., 1975). Galbraith (1994) reached this conclusion by comparing both the latency and spectral characteristics of the FFR recorded with each electrode montage. He reported that responses recorded using the vertical electrode montage had longer latency and different spectral characteristics in comparison to the responses recorded using the horizontal electrode montage.

The longer latency values for the responses obtained using the vertical electrode montage suggested contributions to the response from more central structures in the brainstem due to the longer travel/conduction time needed to reach these structures. Given this conclusion and our interest in the present study in analyzing brainstem encoding of speech sounds, the FFR will be recorded using a vertical electrode montage.

Sampling rate. The sampling rate refers to the number of times a recording system analyzes an analog signal and digitally samples it. Skoe and Kraus (2010) have explained that the sampling rate should be directly related to the Nyquist frequency, that is, the highest frequency in the stimulus. The Nyquist theorem states that a sampling rate should be, at a minimum, twice as large as the highest frequency present in the stimulus, to avoid “gross distortion” of the input signal (Daube & Rubin, 2009). The majority of sampling rates used in FFR studies have emphasized over-sampling (7,000-50,000 Hz) in FFR recordings (Banai, Hornickel, & Skoe. 2009; Krishnan, Xu, Gandour, & Cariani, 2005; Musacchia et al., 2007). Over-sampling is recommended for two reasons, one is to reduce or avoid sampling errors, and the second is for increased temporal precision of the recording to be able to accurately analyze any changes that occur in the signal. Skoe and

Kraus (2010) have suggested employing a sampling rate anywhere between 6,000-20,000 Hz, always taking the Nyquist frequency into account. Based on this recommendation and the stimulus spectrum of our current /u/ stimulus (the highest frequency present in this stimulus occurs at 360 Hz), the present study will record the FFR using a sampling rate between 6,000-20,000 Hz.

Analog band-pass filter setting. An analog filter is a mechanism that eliminates contributions from unwanted electrical or muscular activity (e.g. artifact, electrical interference) from a desired neural signal (Hall, 2007). There are four types of analog filters. These are low pass, high pass, band pass, and band reject filters. Each of these types of analog filters can be defined by their cutoff frequencies and rejection rates or filter skirts. A cutoff frequency is the frequency where the output of filter is reduced by 70.7% of input or 3 dB below down point (Hall, 2007). On the other hand, rejection rate refers to how fast the output is reduced for frequencies that fall outside the pass band (Hall, 2007). A low pass filter has one cutoff frequency where all the energy below that frequency is passed and the energy above the cutoff frequency is rejected or cut. On the other hand, a high pass filter also has one cutoff frequency and only the energy above the cutoff frequency is passed. A band pass filter combines both a low pass and a high pass filter. Therefore, a band pass filter has two cutoff frequencies; the band pass filter passes a band of frequencies located between the two cutoff frequencies. This type of filter is especially useful when attempting to conserve energy present within a defined range, as in the case of the FFR.

The appropriate selection of the analog filter setting is highly important as the underlying energy present in the response can be either enhanced or diminished based on its setting. Thus, selecting appropriate analog filter settings increases the signal-to-noise ratio (SNR) of the response since more energy that is present in the response is captured. Skoe and Kraus (2010) have recommended using an analog band pass filter setting similar to the one used to record a click-ABR (100-3000 Hz) because FFR activity is best reflected between those limits.

Another consideration in selecting an appropriate band pass filter setting is that the auditory brainstem's phase-locking ability (as represented in the FFR) occurs optimally in response to lower frequencies. Glaser, Suter, Dasheiff, and Goldberg (1976) investigated the FFR in humans to both continuous tones and tone burst stimulation. The FFR could be clearly recorded with stimuli within the frequency range of 70-1500 Hz. These investigators suggested that FFR primarily reflects phase-locking to a low frequency stimulus. However, as stimulus frequency increased, the brainstem's phase-locking ability decreased and the FFR eventually ceased (Glaser et al., 1976). Given, that the auditory brainstem's phase-locking ability occurs within this limited low to mid frequency range, the analog band pass filter setting for successfully recording the FFR needs to be set to accommodate this physiological property of the neurons.

Lastly, the ultimate criteria used to select band pass filter settings should also be based on the spectral characteristics of the stimulus. Given the spectral content of the FFR and the stimulus selected for the present study (English vowel /u/ with F0=120 Hz, F1 =360 Hz), an analog filter setting of 100-3000 Hz will be used. Wider band pass filter settings (30-3000) have also been successfully used in recording the FFR (Galbraith, Amaya, & de Rivera, 2004; Galbraith, Arbagey, & Branski, 1995). Ultimately, the selection of appropriate analog band pass filter settings allow for isolation/distinction of subcortical activity from cortical activity.

Number of trials (sweeps). The term *sweeps* refers to the number of stimulus repetitions. As with any electrophysiological response when more sweeps are recorded, a more robust and reliable recording is obtained. This is because the SNR of the response is directly related to the number of sweeps (i.e., the number of sweeps is proportional to the square root of the SNR). Therefore, as you increase the number of sweeps, typically a higher SNR is achieved. Determining an exact number of sweeps to obtain can be difficult as there are many factors to consider (e.g., stimulus type, stimulus intensity, subject factors, stimulus duration, and number of stimuli within an experimental session). When the FFR is recorded to high intensity tones, about 1000-2000 sweeps are needed, similar to the amount click-ABRs require. However, when a stimulus such as speech is used, more sweeps are needed since the speech stimulus is complex and consist of multiple spectral components, thus requiring greater temporal precision for purposes of analysis (Skoe & Kraus, 2010). Skoe and Kraus (2010) have recommended using between 2000-3000 sweeps per stimulus polarity, therefore totaling in the range of 4000-6000 sweeps for alternating polarity stimuli. This recommendation has been made on the premise that if enough sweeps are recorded, response replicability can be determined, and analysis of changes in the response can be made over the course of time. The present study will collect between 2000-3000 sweeps per stimulus polarity with an emphasis placed on response replicability.

Length of the averaging window. Signal averaging in the EP literature refers to the process of summing the responses to repeated stimulus and dividing the summed responses by the number of stimuli presented at discrete time intervals (Hall, 2007). More specifically, the length of the averaging window is defined as the total time period, usually expressed in milliseconds, in which the desired neural signal is averaged. The length of the averaging window can play a significant role in the analysis of the FFR. For an FFR to be optimally recorded, the length of the averaging window should be long enough to accommodate a pre-stimulus baseline period, a response period, and a post-stimulus period. When a pre-stimulus baseline period is included in the length of the averaging window, it also allows for the inclusion of ambient EEG in the recording. This ambient EEG allows for easier interpretation of where the FFR begins in the recording (i.e., amplitude of the FFR must exceed the baseline amplitude). It has been recommended that the pre-stimulus baseline period should be at least 40 ms in length (Skoe & Kraus, 2010). The response period depends directly on the duration of the stimulus. In the present study, the response period will be set at 300 ms since the duration of the stimulus is 250 ms. Lastly, the post-stimulus period should also be included in the calculation of the length of the total averaging window since it helps the audiologist feel confident about the response returning to the baseline. The post-stimulus period should range between 10-50 ms (Skoe & Kraus, 2010). Taken together, the pre-stimulus, response and post-stimulus response periods sum up to a total analysis window length of 300 ms (40 ms+250 ms+10 ms).

Artifact. According to Skoe and Kraus (2010), there are four types of artifacts that can affect FFR recordings. These are electrical noise, myogenic artifact, the (CM), and stimulus artifact. Electrical noise, is usually the result of line noise (50-60 Hz from electric socket) and can be reduced or eliminated by recording the FFR in an electrically-shielded room free of electronic devices and turning on the line filter.

Myogenic artifact refers to muscular artifact (i.e., neck, scalp). Klass (1995) has defined myogenic artifact as being a ‘biologic’ artifact. That is, physical artifact that arises from the subject/patient. Myogenic artifact is worrisome when recording the FFR because of the wide band pass filter settings used in this response. An artifact rejection criterion of +/− 20 μV to +/− 75 μV was suggested since the FFR typically occurs within this range and myogenic artifact exceeds it (Skoe & Kraus, 2010). However, smaller and less obvious muscular artifact may not exceed this range and can contaminate the response. For this reason, subjects should be closely monitored and kept in a relaxed state for the duration of the recording. In the present study subjects will be encouraged to close their eyes and try to sleep to avoid myogenic artifact.

The CM is an electrical potential that occurs within the cochlea. As previously mentioned in the history section on the FFR, the CM, like the FFR, is able to mimic the temporal waveform of an acoustic stimulus. Early investigators who investigated the FFR had a difficult time separating both responses since they share similar characteristics. It is important to understand that even though both responses are able to mimic the temporal properties of an acoustic stimulus, each occurs at a different latency. The CM generally occurs almost instantaneously, whereas the FFR occurs approximately 6-10 ms following the presentation of the stimulus. Furthermore, another distinguishing factor is that the FFR is sensitive to both presentation rate and masking intensity, while the CM is not. In understanding the differences of each response, an audiologist can look at the response waveform and eliminate the CM from the actual FFR.

Lastly, stimulus artifact is an electrical artifact that is produced by acoustic stimulus transducers (Hall, 2007). These acoustic transducers produce electromagnetic fields which in turn result in electrical activity that can muddle the desired electrical recordings (e.g., FFR). Stimulus artifact can be eliminated by using electromagnetically shielded earphones as well as separating electrode leads from the transducer wires (Hall, 2007). The present study will try to control for each type of artifact by recording the FFR in an electrically-shielded room free of electronic devices and turning on the line filter, employing an artifact rejection criterion of +/− 20 μV to +/− 75 μV and encouraging relaxation/sleep.

Subject Parameters

When an AEP is recorded on any living organism, there are specific physiological and anatomical factors and characteristics specific to that organism, which need to be understood and controlled for prior to recording. There are two known subject-related parameters that can affect the ability to record an optimal FFR. These are subject state and age/maturation, both of which are discussed in the section below.

Subject state. Subject state, specifically attention, has been a topic of great debate in the FFR literature. The FFR, like the click-ABR, can be successfully recorded in a sleeping condition (Aiken & Picton, 2008; Dajani, Purcell, Wong, Kunov, & Picton, 2005; Krishnan et al., 2005). Due to the possible influence of myogenic artifact, it is important for the clinician recording the FFR to encourage their subjects to relax (Skoe and Kraus, 2010).

The FFR has also been recorded in active conditions where patients are asked to perform specific tasks such as counting oddball stimulus tokens (Musacchia et al, 2007).

Galbraith and colleagues have performed a multitude of studies which have shown that attention plays a significant role on the FFR (Galbraith & Arroyo, 1993; Galbraith, Bhuta, Choate, Kitahara, & Mullen, 1998; Galbraith & Doan, 1995; Galbraith & Kane, 1993; Galbraith, Olfman, & Huffman, 2003). Specifically, selective attention has been shown to alter/modulate signal processing (of sensory affect pathways) which ultimately affects the FFR since this response serves as a "robust indicator of early auditory neural processing." Based on the known effect of attention on the FFR, participants in the present study will not perform an active task during data collection and will be encouraged to sleep.

Age/maturation effects. The present study will record the FFR in normal hearing adults. However, there are known maturation effects on the FFR. Jeng et al. (2010) investigated early maturation of the FFR to the rising voice pitch stimuli /i/ in infants with normal hearing. In this study Jeng and colleagues made a direct comparison of the FFR acquired in adults to those obtained in the infant group. These investigators reported that the FFR in infants at 2-3 months approximated closely to the FFR acquired in the adult group. Based on these findings Jeng et al. (2010) concluded that the auditory mechanism responsible for the FFR reaches adult like maturation by 2-3 months of age, suggesting that the human auditory brainstem matures during the early stages of life.

In contrast, Johnson, Nicol, Zecker, and Kraus (2008), compared FFRs in two age groups (3-4 years old vs. 5-12 years old) and reported significant differences in "onset synchrony and sustained, phase-locked activity" as reflected in the FFR between the two age groups. Furthermore, these researchers also suggested developmental differences between responses to non-speech stimuli and speech, with the latter being highly dependent on experience and neural plasticity.

Supporting the above findings are results from Clinard, Tremblay, and Krishnan (2010) who have also found that aging alters the perceptual and physiological neural encoding of frequency. Clinard and colleagues did so using the FFR. A total of 32 normal hearing adults between the ages 22-77 years were included in this study. FFRs were recorded using 500 Hz and 1000 Hz tonebursts. Pitch discrimination was assessed using frequency discrimination difference limens (FDLs) and neural representation of these tone bursts was assessed using a linear regression of the FFR phase coherence and FFR amplitude. Both assessments showed a decline in neural encoding and pitch perception ability as their subjects' ages increased. Because only normal hearing adult participants between the ages of 20 and 35 years will be used in the present study, the effect of maturation and aging on the FFR will not be of concern. Aside from the effects of various subject parameters on the FFR, adverse listening conditions have also been found to effect the FFR. The next section will discuss what adverse listening conditions are and their effect on the FFR.

Adverse Listening Conditions

In an ideal listening world, an acoustic signal is received by the auditory system unperturbed; in other words, the integrity of the signal at the level of the ear is identical to that of the output signal of a specific source. However, this is not the case with real world listening experiences; there are many variables in listening environments that can contaminate a signal to the point where it becomes unrecognizable or distorted, resulting in adverse listening conditions. Examples of such adverse or challenging listening situations include listening in background noise and reverberation. As the objective of the present study is to examine the effect of reverberation on subcortical encoding of speech, this section will focus solely on the phenomenon of reverberation and its effect on perception and encoding of speech.

Reverberation. The prolongation of a signal in an enclosed listening environment as a result of its reflection off of hard and non-absorptive surfaces is known as reverberation (Lipscomb, 1978). These surfaces include walls, floors, and ceilings. In any given room where an acoustic signal is produced, there exist two sounds, the direct sound (acoustic signal not yet reflected off any surface) and the reverberant sound (Boothroyd, 2002). Furthermore, sound waves of each type interact with each other and end up distorting and affecting the underlying spectral characteristics of the original signal. If a listener's head is present in the environment, they will receive both sounds. The proportion of direct and indirect sound received depends directly on the proximity of the listener to the sound source. The closer an individual listener is to the sound source, the more direct sound they will receive. However, the further away the listener is to the sound source, the less direct sound and more reverberant sound they will receive, since they are closer to the reflective surfaces than the source itself. The mid-point location of a listener in an environment where they receive an equal amount of reverberant sound and direct sound is known as the critical distance (as illustrated in figure 18 below).

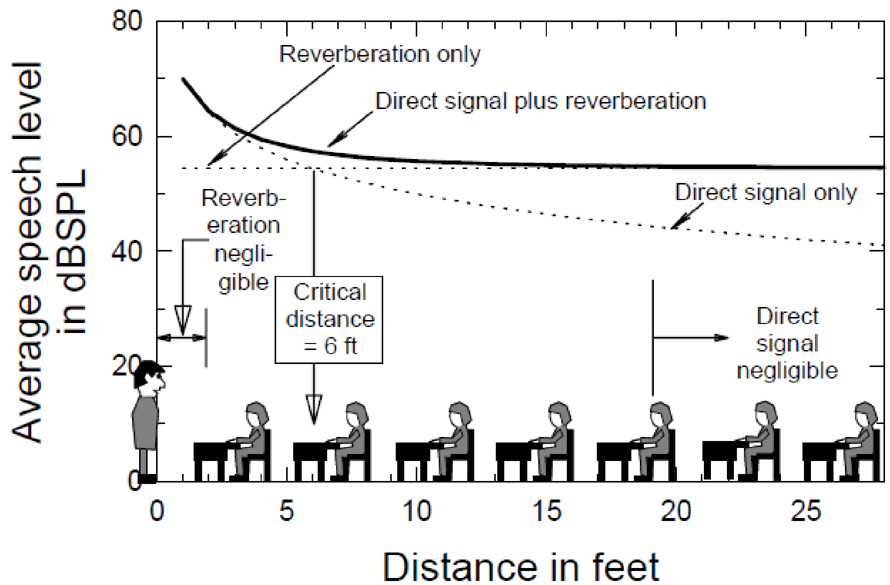

Figure 2. Predicted long-term average speech level as a function of distance from the talker in a room measuring 30x20x9 feet with a reverberation time of 0.5 seconds.

Figure 17. Adapted from "Room Acoustics and Speech Perception," by A. Boothroyd, 2002, *In Seminars in Hearing*, 25, p. 155-166. Copyright 2002 by Thieme Medical Publishers Inc.

Controlling reverberation can be a challenging task; it requires a balance of both reflective and absorptive surface materials. Reflective materials enhance reverberation whereas absorptive materials control or weaken reverberation in listening environments. An ideal listening environment should be balanced to have just the right amount of reverberation. According to Lipscomb, enclosed listening environments which lack sound reflections are referred to as "dead" rooms (e.g., anechoic chambers) and those which have some degree of reflection are referred to as "live" rooms. Specifically, Lipscomb (1978) stated that a "certain degree" of reverberation is acceptable and can make for a preferable listening environment. The degree of reverberation present in an environment is defined by a measure known as reverberation time (RT). RT is the amount of time it takes for a signal and its components, to decay by 60 dB, after its production by a source has ceased (Crandell & Smaldino, 2000). Thus, a higher degree of RT represents increased levels of reflection, whereas a lower RT indicates a decreased reflective environment. An acceptable degree of reverberation is less than 1 second of reverberation time, and ideally between 0.4-0.6 RT (Hodgson & Nosal, 2001).

The interaction between the original and reflected sound waves that occurs in reverberation “smears” the temporal characteristics of the original sound. As a result, when there is a higher degree of reverberation present in a listening environment, the acoustic make-up of speech signals is affected, leading to deficits in speech perception. It is important to understand how different degrees of reverberation acoustically affect a listening environment and ultimately, speech perception.

Reverberation and speech perception. The behavioral effects of reverberation on speech perception have been investigated extensively. Kreisman (2003) explained that reverberation effects on speech perception (e.g., word recognition score testing) differ between normal hearing individuals and those with sensorineural hearing loss (SNHL). For example, in normal hearing individuals, speech perception abilities do not begin to decline until RT is greater than one second. On the other hand, if an individual has a SNHL, speech perception abilities become compromised sooner, when RT is greater than 0.4-0.5 seconds. Many studies have used speech identification tasks to study the effects of reverberation on speech perception in normal hearing individuals and listeners with SNHL. Stimuli in these investigations have included sentences, words, consonants, diphthongs, and vowels (Febo, 2003; Hazrati & Loizou, 2012; Hodoshima & Arai, 2007; Humes & Roberts, 1990; Nabelek, 1988; Nabelek & Dagenais, 1986; Nabelek, Letowski & Tucker, 1989; Nabelek & Robinson, 1982). Since a vowel sound will be used in the present study, this section will focus primarily on behavioral studies that have investigated speech perception in reverberant conditions using shorter duration stimuli such as consonants, diphthongs, and vowels. Investigations that have employed diphthongs and consonants are included in this review since there are a very limited number of studies that have exclusively used vowels as a stimulus.

General effects of reverberation. Nabelek and colleagues published a series of studies on the effects of reverberation on the speech perception of words and short duration stimuli (e.g., consonants and vowels). The first of many studies in this area focused specifically on monaural versus binaural speech perception abilities in reverberant conditions amongst various age groups. In this study, Nabelek and Robinson (1982) presented words from the Modified Rhyme Test (MRT) both monaurally and binaurally at three different RT levels (0.4 sec, 0.8 sec, and 1.2 sec).

Researchers found that although a level of 70 dB SPL was optimal for speech perception across all age groups, they reported that the young adult group was able to achieve optimal performance at 50-60 dB SPL. In contrast, the young children and elderly groups required higher presentation levels to achieve optimal performance. Nabelek and Robinson concluded that as the subjects' age increased, a higher intensity level was needed to achieve performance scores similar to the young adult group. These investigators also reported that a binaural advantage was demonstrated in all reverberant conditions for all age groups tested. Although all age groups benefitted from binaural listening on speech perception tasks, the group that saw most benefit were the elderly groups (5.0-6.7% increase versus 3.8% increase in young adult group.) Results from this investigation signify that listening with two ears versus one in reverberant conditions markedly improves the subjects' speech perception abilities.

Effect of reverberation on short segment stimuli. Four years later, Nabelek and Dagenais (1986) investigated vowel errors in reverberation by hearing impaired listeners. These vowels could be sub-grouped into monophthongs (e.g., /i/, /o/, /ae/, /u/) and diphthongs (e.g., /au/, /aɪ/). In the preliminary stages of this study, normal hearing subjects were also tested using a vowel identification test presented with a RT of 1.2. All normal hearing participants scored in the range of 97%-100% on identification tasks. In contrast these monophthongs and diphthongs were highly misidentified by hearing impaired listeners (mean score of 76.8%). These results indicate that speech understanding is impacted to a greater extent in individuals with SNHL in reverberant listening conditions. Specifically, Nabelek and Daegenais (1986) noted that the listeners had difficulties differentiating between the shorter monophthong and longer diphthong stimuli, since in reverberant conditions the prolongation of monophthongs can be misperceived as the longer duration diphthongs, because of temporal smearing. The figure below illustrates spectrograms of the monophthong /ʌ/ and the diphthong /aɪ/, both of which show temporal smearing in the reverberant condition versus the quiet condition. Visually, the spectral components of each vowel are easier to distinguish in the quiet condition than in the reverberant condition. In the spectrogram recorded in reverberation, there is a prolongation of all components that makes it more difficult to separate individual components.

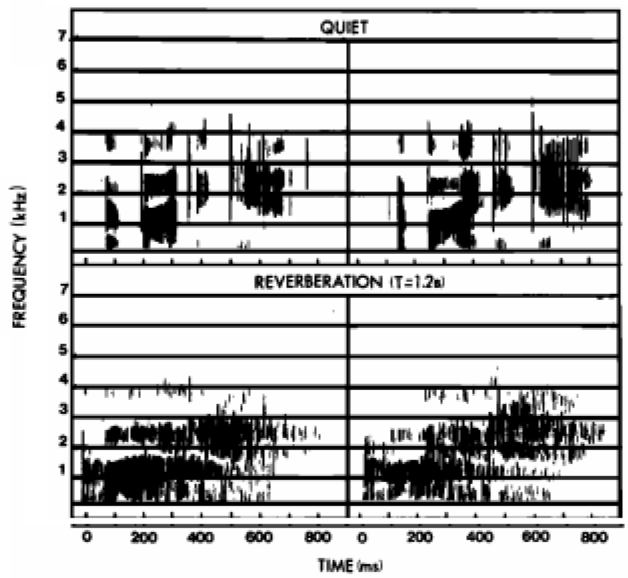

FIG. 1. Spectrograms of the phrases "(th)e/bʌt/ again" and "(th) e /bɑɪt/ again" in quiet, ;, and reverberation.

Figure 18. Adapted from "Vowel Errors in Noise and in Reverberation by Hearing-impaired Listeners." by A.K. Nábělek and P.A. Dagenais, 1986, *The Journal of the Acoustical Society of America, 80 (3)*, p. 741-8.

Thereafter, Nabelek et al. (1989) investigated two known effects of reverberation (overlap-masking and self-masking) on consonant identification. According to Nabelek (1989) overlap-masking is an effect that results when the spectral energy of preceding consonant contaminates that of the following consonant. Thus, the preceding consonant has a masking effect on the consonant that follows. On the other hand, self-masking refers to the "internal temporal smearing of energy within each consonant." Self-masking means that a phoneme is not affected by anything other than itself (e.g., phoneme in initial syllable position).

The ultimate goals of the Nabelek et al. (1989) study were to measure how much self-masking and overlap masking affect identification in reverberation respectively, and to examine if the effects of reverberation on speech perception are similar to effects of various types of background noise. A total of eight consonants were used in this study (e.g., /p, t, k, f, m, n, l, w/), each of which was preceded by the consonant /s/, and followed by the phoneme, /at/ (i.e, /spat/, /stat/, /swat/). Each speech stimulus was presented to a total of ten normal hearing participants (mean age=32.3 years) in four conditions: in quiet, with a RT of 1.9 sec, with a babble masker (S/N= +5 dB), and with a /s/ noise masker set at a signal to noise ratio of +3 dB. Nabelek and colleagues reported that the effects of reverberation alone on consonant identification were separate to those noted in the two other noise conditions. These investigators found that there were three identification errors specific to the reverberation alone test condition. The first error confirmed that overlap masking was caused by the preceding phoneme in a stimulus (i.e., reverberation caused more errors for consonants in the /s-at/ [191 errors] than in the /-at/ [65 errors] context). The second type of error in reverberation was that of /-at/ context errors which could be directly attributed to self-masking since these errors occurred in the initial position, this type of error was not present in background noise. The last and third error was related to the masking of co-articulation cues between the preceding /s/ stimulus and the consonants.

These two known effects of reverberation on identification are significant, and they must be understood, especially for purposes of analyzing the how reverberation affects subcortical neural encoding, as in the case of FFR.

Neural encoding and reverberation. To date, there is limited literature on the effects of reverberation on human neural encoding. In 2008, Sayles and Winter performed an experiment to explore how reverberation effects the temporal representation of the pitch of complex sounds in animals. These investigators examined this by recording action potentials at the level of ventral cochlear nucleus (VCN) in guinea pigs. Recordings were taken of 240 isolated single-units in the VCN of these guinea pigs using F0-swept harmonic complexes. According to Sayles and Winter (2008) different single-units within the VCN are known to respond differently to acoustic stimulation. Generally speaking, some, as in the case of primary-like (PL) units respond similarly to auditory nerve fiber input, in that they preserve the accurate timing of information that is necessary for precise encoding of the stimulus temporal fine structure. On the other hand, chopper units produce much more complex responses. These have poor phase-locking abilities to temporal fine structure but enhanced ability with temporal-envelope modulation.

The first part of the study recorded neuronal responses to dynamic pitch stimuli with added reverberation. Researchers in this segment of the study measured the neuronal responses by analyzing the temporal pattern of the action potentials and adding the interspike-interval distribution. Based on these two measurements a calculation of the pitch perception was made. The dry condition produced the best pitch period but pitch period greatly decreased in reverberation.

Researchers Sayles and Winter also found that the biggest effect of reverberation on single-unit responses occurred in PL units. In short, neurons tuned to lower frequencies were able to maintain a robust “representation of the stimulus fine structure”, specifically the time-varying F0, even in the presence of high levels of reverberation but those tuned to higher frequencies were not.

Phase-locking ability in reverberant conditions was also examined. There was a significant difference in phase-locking abilities between different units in the VCN. PL units were able to phase-lock to frequencies below 3500 Hz in a dry condition, whilst chopper units were limited to 1500 Hz and below. Reverberation appeared to have a negative effect on the pitch-related response, even in mild levels. Although, some units (PL) could phase-lock to frequencies upwards of 3500 Hz, these did not seem to provide any significant increase in temporal representation to pitch in the harshest reverberation condition. Regardless of the range of phase-locking abilities across different VCN single-units, all of these showed markedly robust temporal pitch representation in response to low frequencies, regardless of condition.

As previously mentioned in the auditory neural encoding section in the basic acoustics chapter, neural encoding to auditory stimuli involves two processes, the slowly-varying E and the TFS. The E has been suggested as a cue, helpful in figuring out the pitch of a complex sound; however, as reverberation is increased, the effectiveness of this cue decreases. In reverberation, TFS is much more affected than the E, and although reverberation decreases E cues that are helpful in determining pitch, Sayles and Winter (2008) have suggested that phase-locking to TFS is more essential to increased behavioral performance in humans. Given that TFS cues play an important role in adverse listening conditions, and the finding that phase-locking to TFS cues is diminished in reverberant settings may explain why normal-hearing listeners struggle in such adverse listening conditions.

Further, behavioral studies have established that listeners with SNHL rely more on E cues than TFS cues, as TFS representation is degraded to begin with in SNHL. When these crucial E cues are decreased or absent in the presence of reverberation, the auditory system is left with trying to use temporally smeared TFS information, which is already compromised, to resolve the missing auditory information, making pitch perception extremely difficult. Hence, listeners with SNHL experience greater difficulty with speech perception in adverse listening conditions.

Reverberation and the FFR. Although the effects of reverberation on speech perception have been studied extensively, few studies have examined the effects of reverberation on neural encoding of speech. Researchers Bidelman and Krishnan (2010) investigated the effects of reverberation on brainstem neural encoding of a vowel sound in ten musicians and ten non-musicians with normal hearing using the FFR. The stimulus used was a synthetic version of the vowel /i/ with a rising time-varying F0 contour that increased from 103-130 Hz over the course of 250 ms. However, the formant frequencies of this stimulus were steady state (e.g., F1= 300 Hz, F2=2500 Hz, F3=3500, F4=4530). This vowel stimulus was presented in three different RTs (mild reverberation (0.7 sec), medium reverberation (0.8 sec), and severe reverberation (0.9 sec). As a control, a “dry” vowel stimulus with no modification or reverberation was also presented. FFRs were recorded monaurally at 80 dB SPL using a rarefaction stimulus polarity and a repetition rate of 2.76 per second.

Investigators noted three observations on the effect of reverberation on the FFR. First, as reverberation levels were increased, the temporal resolution of the formant-related harmonics decreased. However, the F0 appeared to maintain/preserve its integrity regardless of reverberation levels. This preservation can be noted in figure below. On the left side of this figure, a spectrogram with increased temporal resolution can be visualized. In the no reverberation (A) or “dry” condition multiple horizontal striations can be seen. The lowest of these represents the fundamental frequency; each striation superimposed on top of this one is a formant-related harmonic. Both components are clear in the dry condition, but as reverberation levels are increased (B-D), the clarity of these striations (spectral components) decreases. On the right side of the figure is a spectrogram with increased frequency resolution (E). The F0 is preserved in all four test conditions as an evident peak can be noted at each level. Although the peaks of each formant related harmonic are clearly visualized in the dry condition, the ability to differentiate between each peak as the degree of reverberation was increased declined, resulting in smearing of higher frequency components.

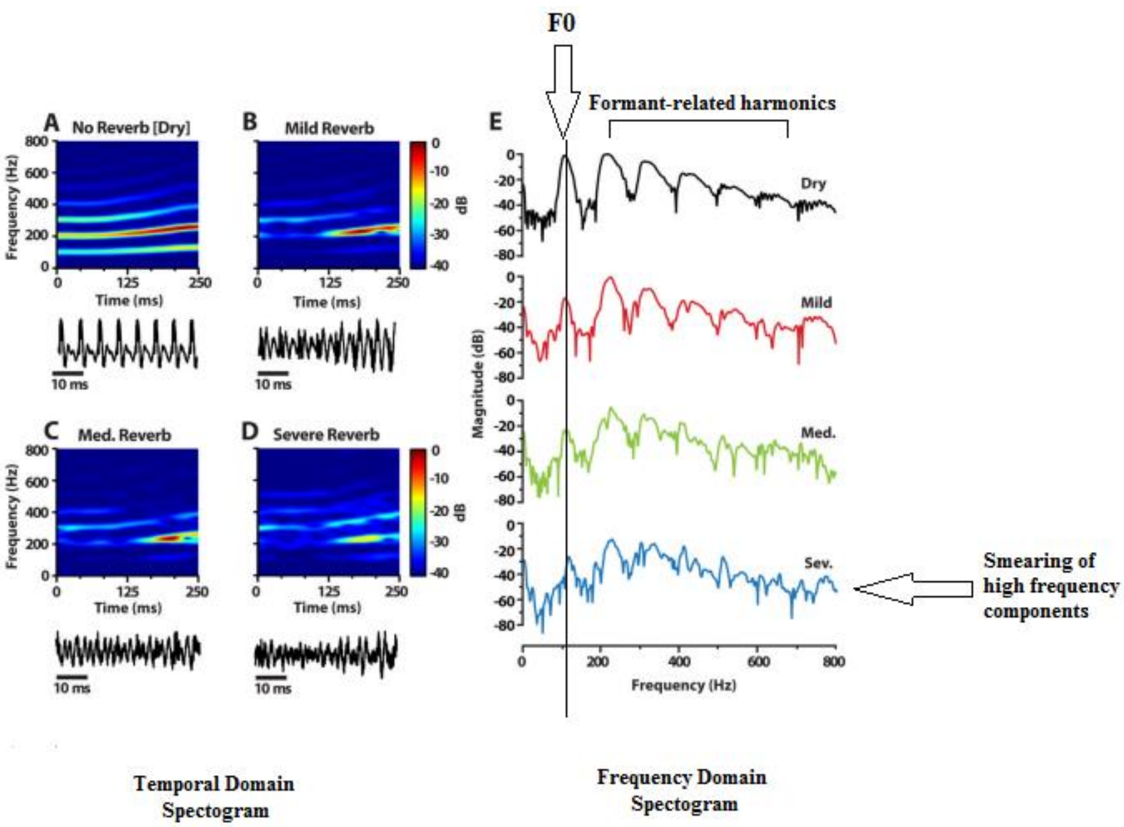

Figure 19. Temporal domain spectrograms (left side) and frequency domain spectrograms (right side) of FFRs recorded to varying levels of reverberation (dry, mild, medium, and severe). Adapted from “Effects of Reverberation on Brainstem Representation of Speech in Musicians and Non-musicians,” by G.M. Bidelman and A. Krishnan, 2010, *Brain Research*, *1355*, 12-125.

Second, the brainstem encoding of F0 and F1 components was more pronounced in musicians than non-musicians in all conditions, indicating that musical experience can enhance subcortical neural encoding. Not only that, but grand average FFR temporal waveforms showed that musicians had salient and more robust response periodicity (can be visualized in figure 21 below).

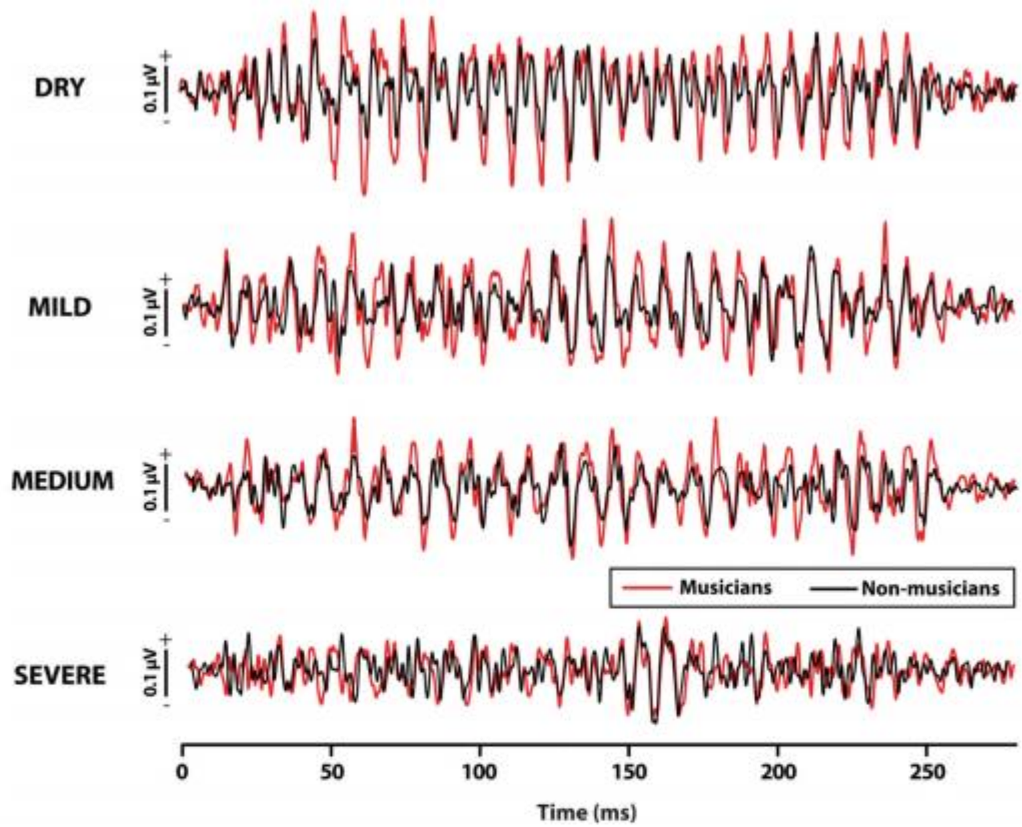

Figure 20. Grand average FFR temporal waveforms obtained in musicians (red) and non-musicians (black) to varying degrees of reverberation (dry, mild, medium, and severe). Adapted from "Effects of Reverberation on Brainstem Representation of Speech in Musicians and Non-musicians," by G.M. Bidelman and A. Krishnan, 2010, *Brain Research*, *1355*, 12-125.

Lastly, researchers in this study also performed a psychophysical experiment to complement the FFR study. In this part of the study, musicians were able to discriminate between the F0 and the first formant-component harmonic, with 4 times the acuity observed in non-musicians. Therefore, the differences noted in the FFR results were also reflected in the behavioral study. This similarity in differences suggests that the FFR is an effective physiologic tool for examining effects of reverberation on neural encoding. Based on these findings, and taking into account that most individuals are not musically experienced, overall these results show that the primary effects of reverberation on the FFR are on encoding of high-frequency formant related harmonics and decreased temporal precision (phase-locking) in normal hearing individuals as reverberation levels increase.

On the whole, the present discussion has reviewed the effects of reverberation on three different responses: behavioral responses, the FFR in humans, and action potentials at the level of the VCN in guinea pigs. Behaviorally, increased reverberation levels have been shown to decrease performance on word identification tasks. In the FFR, as evidenced in the Bidelman and Krishnan (2010) study, increased reverberation levels resulted in decreased magnitudes of formant-related harmonics as well as of temporal precision. Increased reverberation in Sayles and Winter (2008) study resulted in degradation of E cues and phase-locking ability. The effect of reverberation on perception is not to be underestimated; negative effects reported across these different studies confirms that more research is needed to understand human auditory neural encoding in this condition.

Goals/aims of the Current Study

1) To determine the effect of reverberation on neural encoding of F0 and F1 components in normal hearing individuals using the FFR.

2) To gain further insight into the listening difficulty of normal hearing individuals in adverse listening conditions.

CHAPTER 3: METHODOLOGY

Participants

A total of six normal hearing young adults between the ages of 23-25 years (M=23.83 years, S.D. = 0.75 years, males= 3, females= 3) participated in the study. Otoscopic examination was performed to ensure clear ear canals. All participants were required to have normal middle ear function which was confirmed using the following tympanometric criteria: peak pressure values between +/- 149 daPa, static compliance values between 0.3 and 1.4 ml, and ear canal volumes between 0.5 and 1.6 ml (Jerger, 1970). Pure-tone thresholds were evaluated at frequencies 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz, bilaterally. A minimum threshold of 15 dB HL was needed at each stimulus frequency in order to be included in the study. Participation in the study was voluntary and all participants were required to give informed consent in compliance with the Institutional Review Board of Towson University.

Stimuli

A synthetic steady-state vowel /u/ (F0=120 Hz, F1=360 Hz) was used as the primary stimulus in this study. The stimulus was generated by *Praat*, a computer software package for analysis of speech. The reverberation was introduced into the stimulus using a *MATLAB* code. The stimulus was then imported into the Intelligent Hearing Systems SmartEP platform and calibrated using a sound level meter. The speech stimulus was presented in four different conditions: no reverberation or dry condition, mild reverberation (RT= 0.6 sec), moderate reverberation (RT=0.8 sec), and severe reverberation (RT=1.1 sec). FFRs were recorded from each participant in response to monaural stimulation of their right ear at a stimulus intensity of 75 dB SPL using magnetically shielded Etymotic ER-3A insert earphones. Each stimulus was presented using an alternating polarity (for easier analysis of TFS and E components) at a presentation rate of 3.13/s.

FFR Data Acquisition

Participants were asked to relax while seated on a comfortable recliner situated in a sound booth during the recording session. Participants were instructed to keep their eyes closed and were encouraged to sleep. They were also told to minimize movement during recording in order to control for myogenic artifact. All recordings took place in an acoustically and electrically shielded booth to prevent any other artifact during recordings. The presentation order of the different conditions was randomized within participants to avoid any order effects in the responses.

FFRs were recorded in the IHS SmartEP platform using a vertical montage with a non-inverting (+) electrode placed at the midline of the forehead at the hairline (Fz). Inverting (−) reference electrodes were placed on the left and right mastoid (M1, M2) and on the seventh cervical vertebrae (C7). Another electrode was placed on the mid-forehead (Fpz) and served as the common ground. All inter-electrode impedances were monitored and maintained below 1 kΩ. A minimum sampling rate of 6,000 Hz was used. The analog band-pass filter was set from 100-3000 Hz. A total of 2000 sweeps was required per condition. The length of the averaging window was set at 300 ms. Only sweeps containing activity within the ± 20 μV to ±75 μV range were accepted; anything outside this range was rejected as artifact. The length of testing was approximately 2.5-3 hours long. Participants were required to commit to a total of two test sessions.

FFR Data Analyses

Analysis of FFR data was divided into two types, a qualitative data analysis and quantitative data analysis. All FFR data was exported from the IHS platform to MATLAB for the temporal and spectral domain analyses. Temporal waveforms were obtained for each participant per condition. The waveforms were averaged across participants for each condition to yield a grand averaged temporal waveform per condition. These grand averaged waveforms were calculated by either subtracting or adding FFR waveforms recorded to either condensation or rarefaction stimulus polarities.

Qualitative data analysis. Qualitative indices of the neural response at each RT were provided through a visual analysis of the periodicity and root mean square (RMS) amplitude of the grand averaged waveforms obtained at each RT.

Quantitative data analysis. The MATLAB software was used to analyze the temporal waveforms. As previously mentioned, the FFR is a response that occurs in response to complex stimuli. Thus, this response is composed of multiple frequency components, as it mimics the complex sounds it responds to. As the interest of the present study was to analyze the F0 and F1 components, it was necessary to separate these components. In order to separate these components and identify which frequencies have maximum energy (expected at F0 and F1 components), all temporal waveforms were converted into the frequency domain using an FFT analysis. The magnitudes of both the F0 and F1 components for all participants in each condition were measured.

Statistical analysis. Because six participants were tested, inferential statistical methods were not performed. Instead, the actual statistical analysis for this study consisted of descriptive statistics. The mean and standard deviation of response amplitudes (F0 and F1 components) was calculated for all six participants by condition.

CHAPTER 4: RESULTS

The results section is divided into two major sections focusing on: 1) brainstem neural encoding to the F0 (E) component and 2) brainstem neural encoding to the F1 (TFS) component. Within each of these major sections, results from the FFR temporal waveform qualitative analysis are described first followed by FFR spectral analyses (FFT analysis of F0 and F1 and grand-averaged FFT data). This same organization structure will be applied in the discussion section.

$\mathbf{FFR_{ENV}}$ (F0) Temporal Waveform Composition

The grand-averaged FFR temporal waveforms reflecting the summed condition or envelope encoding (FFR_{ENV}) of six normal hearing subjects across four reverberation conditions are shown in Figure 22. Specifically, in order from top to bottom, this figure illustrates the response waveforms obtained to a clean condition (panel A), mild reverberation condition (panel B), moderate reverberation condition (panel C), and severe reverberation condition (panel D) to the /u/ stimulus. Each figure panel shows a temporal waveform of approximately 250 ms in duration. The y- and x-axes represent amplitude (mV) and time (ms), respectively. The organization of panels used in this section will be used for the entirety of the results section.

Visual inspection of the grand-averaged FFR_{ENV} temporal waveforms indicates several interesting trends: (1) a decrease in waveform amplitude with increasing levels of reverberation, (2) a decrease in temporal resolution of waveforms with increasing levels of reverberation, and (3) and the highest degradation in temporal waveform amplitude and resolution in the most severe reverberant condition. The most robust amplitude was noted for the clean condition (~.28 mV) in comparison to other test conditions. The amplitude of the temporal waveform appears to decrease prominently in the mild reverberation condition (~0.20 mV) and remains visibly decreased for the moderate reverberation condition (~0.20 mV). Most notably, the severe reverberation condition shows the least amplitude (~0.15 mV). The amplitude in the severe condition is approximately half of that for the clean condition. Lastly, there are also changes in waveform morphology as the severity of reverberation is increased. Specifically, as the severity of reverberation increases, the temporal waveforms are less resolved with decreased periodicity, indicating decreased neural phase-locking ability. The discrete and salient peaks noted in the clean condition become noticeably blurred across the reverberation conditions, especially in the poorest condition (panel D).

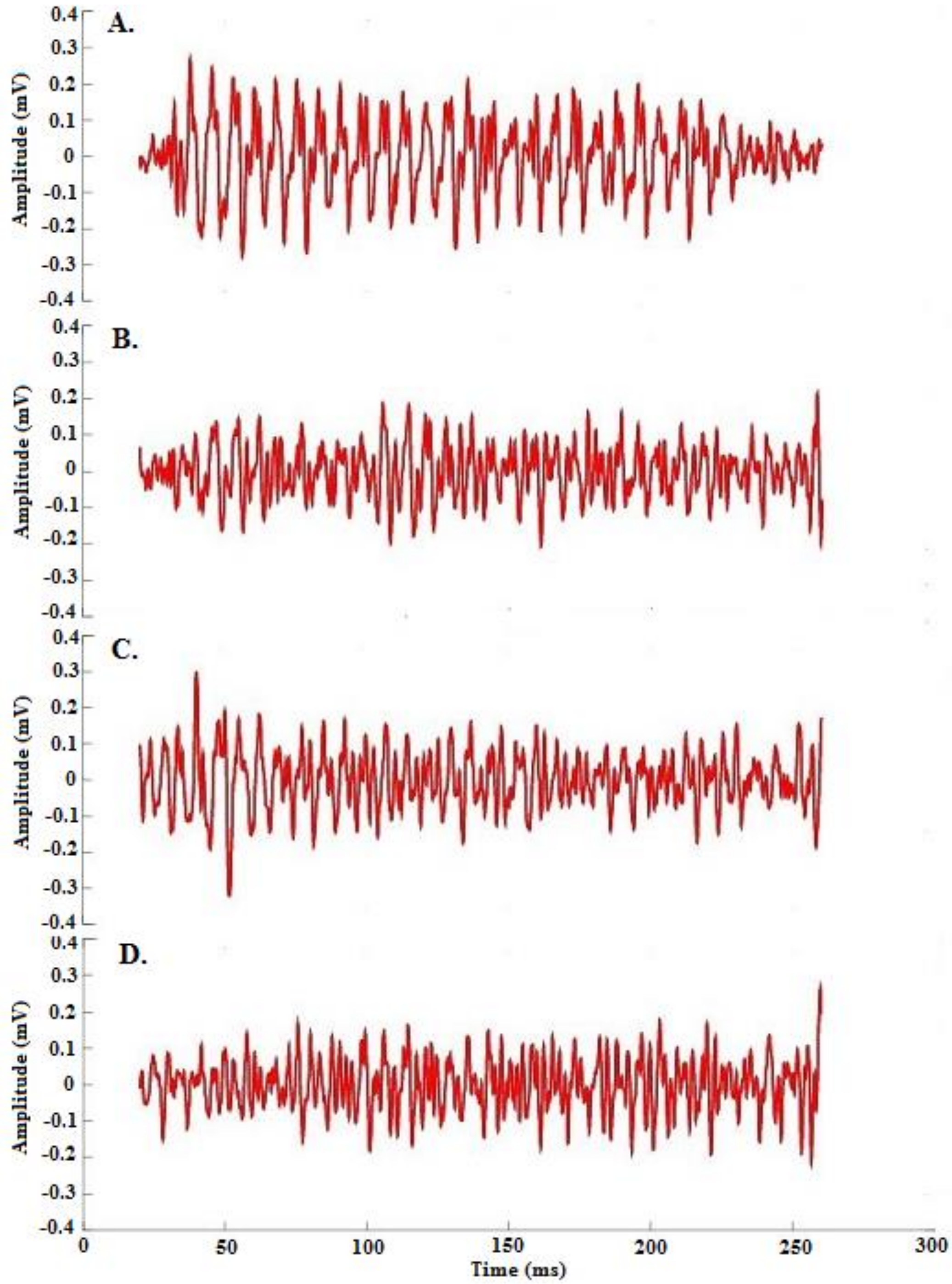

Figure 21. Grand-averaged FFR_{ENV} temporal waveforms across four reverberation conditions. (From top to bottom: Clean Condition (A), Mild Reverberation (B), Moderate Reverberation (C), and Severe Reverberation (D)).

FFR_{ENV} (F0) Spectral Composition

FFT analysis was used to convert FFR_{ENV} temporal waveforms into the frequency domain for spectral analysis. The grand-averaged FFT results reflecting the summed condition or FFR_{ENV} obtained across six normal hearing subjects for each of the four reverberation conditions are shown in Figure 23. Each figure panel shows an FFT with a peak at approximately 120 Hz (which represents the response to the F0 of the stimulus) and the harmonics which occur at whole number integers of the F0. The y- and x-axes represent amplitude (mV) and frequency (ms), respectively.

Observation of FFT spectral data reveals three key findings: (1) spectral energy at the F0, (2) spectral energy at harmonics of the F0, and (3) a substantial decrease in brainstem encoding of the F0 in the severe condition. More specifically, of the four reverberation conditions, the strongest F0 amplitude is seen in the clean condition (0.141 mV). There is a decrease in F0 amplitude as a function of reverberation severity; the peak corresponding to the F0 gradually decreases in amplitude with increasing reverberation (B (0.096 mV)-D (0.075 mV)). The smallest peak at the F0 is shown in the severe reverberation condition seen in panel D. Additionally, as expected, it was observed that the amplitude of the peak at F0 was more robust than any other harmonics in the clean condition. This observation can also be seen for the mild and moderate reverberation conditions. However, it becomes increasingly difficult to distinguish the F0 amplitude peak from the harmonics in the severe reverberation condition.

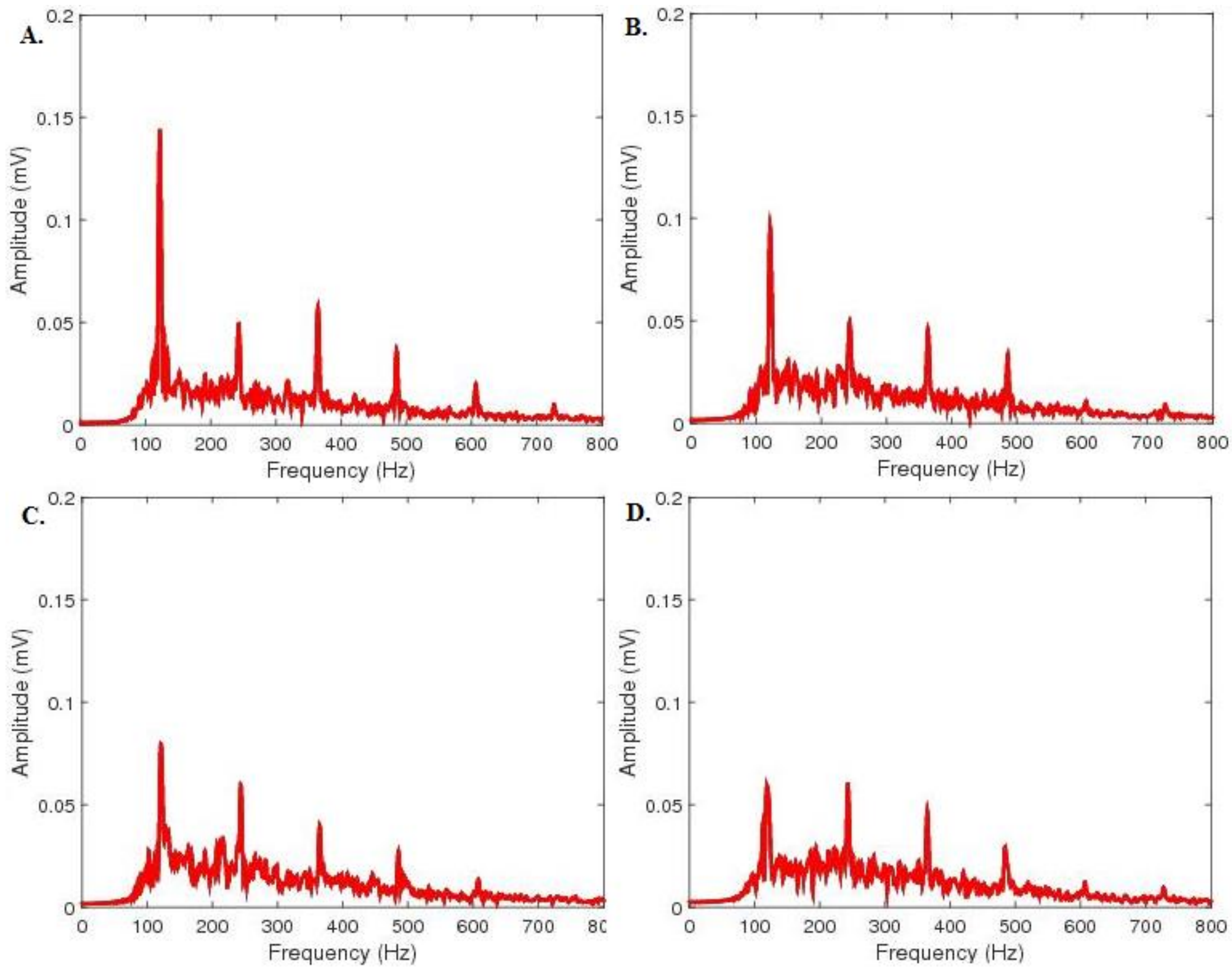

Figure 22. Grand-averaged FFR_{ENV} FFTs across four reverberation conditions. (Panel A: Clean Condition, Panel B: Mild Reverberation, Panel C: Moderate Reverberation, Panel D: Severe Reverberation).

F0 Quantitative Spectral Analysis: Individual and Mean Values

Amplitude values at the F0 were obtained for each test condition. Table 3 presents these F0 values as well as the mean and standard deviation values.

Table 3

Individual F0 Component Amplitudes of Six Normal Hearing Participants for Varying Levels of Reverberation with Descriptive Measures.

	F0 Component				
Participant	Clean	Mild RT: 0.6	Moderate RT: 0.8	Severe RT: 1.1	Difference
S1	0.127 mV	0.073 mV	0.076 mV	0.052 mV	0.075 mV
S2	0.123 mV	0.162 mV	0.083 mV	0.093 mV	0.03 mV
S3	0.118 mV	0.077 mV	0.062 mV	0.063 mV	0.055 mV
S4	0.131 mV	0.081 mV	0.080 mV	0.056 mV	0.075 mV
S5	0.163 mV	0.072 mV	0.121 mV	0.122 mV	0.041 mV
S6	0.182 mV	0.109 mV	0.077 mV	0.065 mV	0.117 mV
M	**0.141** m**V**	**0.096** m**V**	**0.083** m**V**	**0.075** m**V**	**0.066 mV**
SD	**0.025** m**V**	**0.035** m**V**	**0.019** m**V**	**0.026** m**V**	**0.001 mV**

Note. Difference= difference between response amplitude of clean condition and severe reverberation condition.

Several interesting findings were noted in this quantitative data. First, as the severity of reverberation increased, there was a substantial decrease in F0 mean amplitude values, with the smallest amplitude measurement being seen in the severe reverberation condition (.075 mV). Specifically, the mean data showed that F0 amplitude decreased by approximately 50% from the clean condition to the severe condition (.141 mV vs. .075 mV, respectively). The majority of the individual subjects followed this same pattern of a decrease in F0 amplitude values as the reverberation condition worsened. The only exception to this pattern occurred for participants 2 and 5, who showed either minimal or relatively small changes in F0 amplitude values from the quiet to most severe reverberation conditions, as reflected in their difference values.

The variance at each test condition was also calculated as reflected in the standard deviation values. In general, the variance was relatively similar across test conditions, and therefore did not contribute to the mean F0 amplitude differences seen across conditions. A third interesting finding was that the greatest decrease in F0 mean amplitude between sequential conditions occurred between the clean (.141 mV) and mild (0.096 mV) reverberation conditions; whereas the smallest decrease in F0 mean amplitude between sequential reverberation condition occurred between the moderate (.083 mV) and severe (.075 mV) reverberation.

The trends described in this section can be visualized in Figure 24. This figure plots the mean F0 amplitude obtained in each of the four conditions. The y- and x- axes in this figure represent mean amplitude in millivolts (mV) and reverberation condition, respectively.

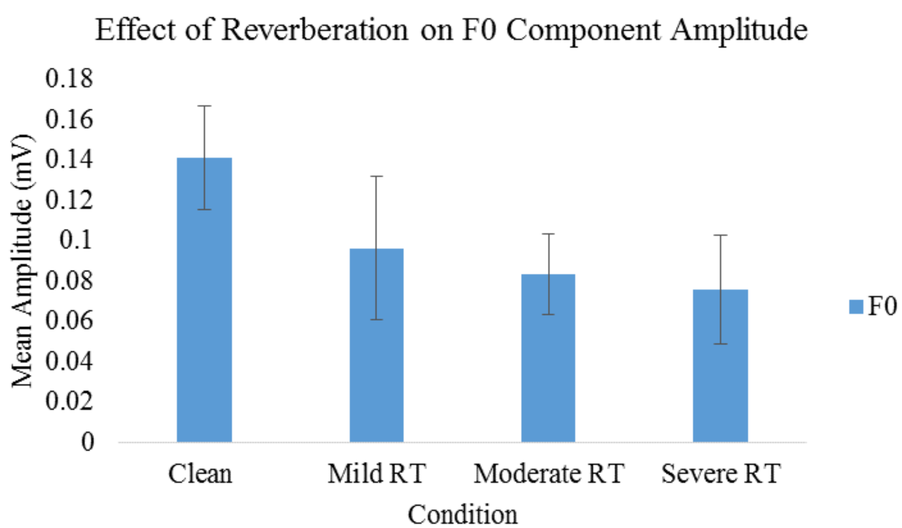

Figure 23. Grand-averaged mean F0 amplitudes of six normal hearing participants across varying levels of reverberation. As reverberation severity increases, mean F0 amplitude decreases. Note the standard deviation bars.

FFR_{TFS} (F1) Temporal Composition

Grand-averaged FFR waveforms reflecting the subtracted condition or fine structure encoding (FFR_{TFS}) across six normal hearing participants for four reverberation conditions are shown in Figure 25. Again, each figure panel shows a temporal waveform which is approximately 250 ms in duration. The y- and x-axes represent amplitude (mV) and time (ms), respectively.

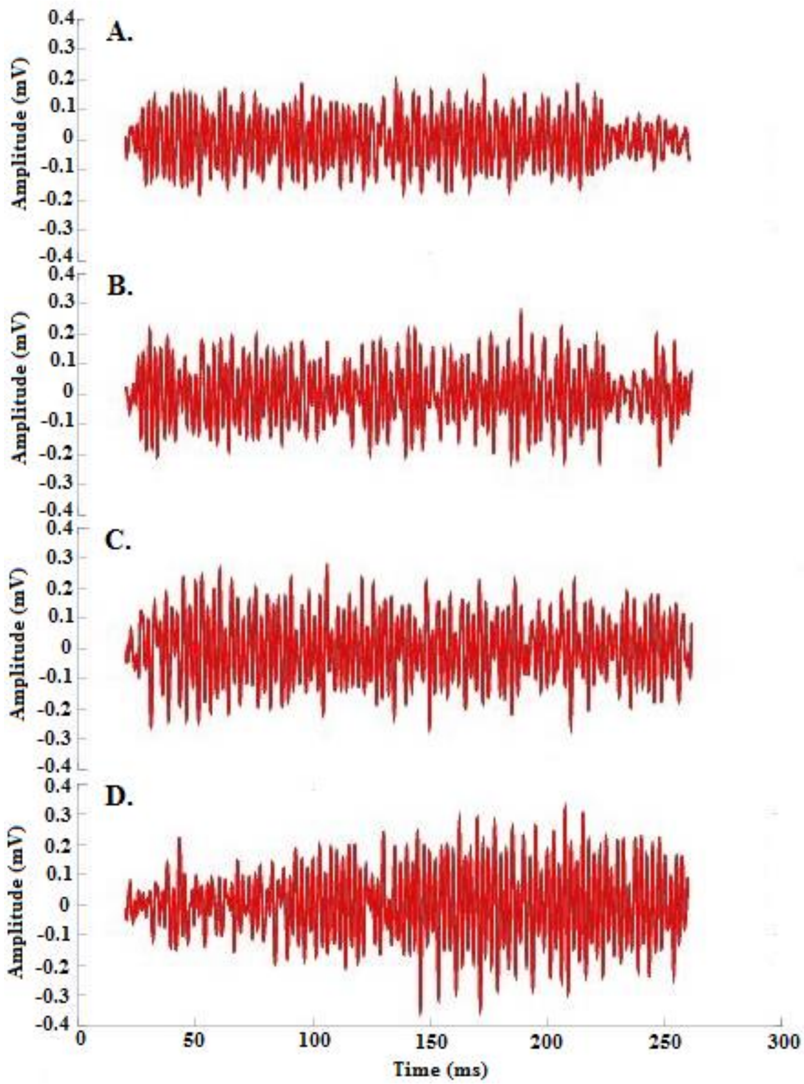

Figure 24. Grand-averaged FFR_{TFS} temporal waveforms of six normal hearing subjects across four reverberation conditions. (From top to bottom: Clean Condition (A), Mild Reverberation (B), Moderate Reverberation (C), and Severe Reverberation (D)).

Visual inspection of the grand-averaged FFR_{TFS} temporal waveforms revealed several findings: (1) the severe reverberation condition showed notable change in waveform morphology, but this was only observed in the first 150 ms of the temporal response. This severe condition shows a sudden decrease in waveform amplitude in the first 150 ms of the response duration (thereafter, the amplitude for the severe condition increases from approximately 0.10 mV to 0.35 mV and remains visibly robust through 250 ms); (2) all waveforms, regardless of reverberation severity, demonstrate a lack of temporal resolution or discrete periodicity; and (3) an unexpected increase in temporal waveform amplitude as a function of reverberation severity (~0.20 mV vs. ~0.35 mV, for the clean and severe conditions, respectively) (apart from the first 150 ms in the severe reverberation condition described above). In particular, the amplitude of the temporal waveforms appears to the smallest for the clean condition (panel A).

FFR_{TFS} (F1) Spectral Composition

FFT analysis was used to convert FFR_{TFS} temporal waveforms into the frequency domain for spectral analysis. The grand-averaged FFTs reflecting the subtracted condition or FFR_{TFS} obtained across six normal hearing participants for each of the four reverberation conditions are shown in Figure 26. Here, each figure panel shows an FFT with a peak at approximately 360 Hz (which represents the response to the F1 of the stimulus) and at the formant-related harmonics. The y- and x-axes represent amplitude (mV) and frequency (ms), respectively.

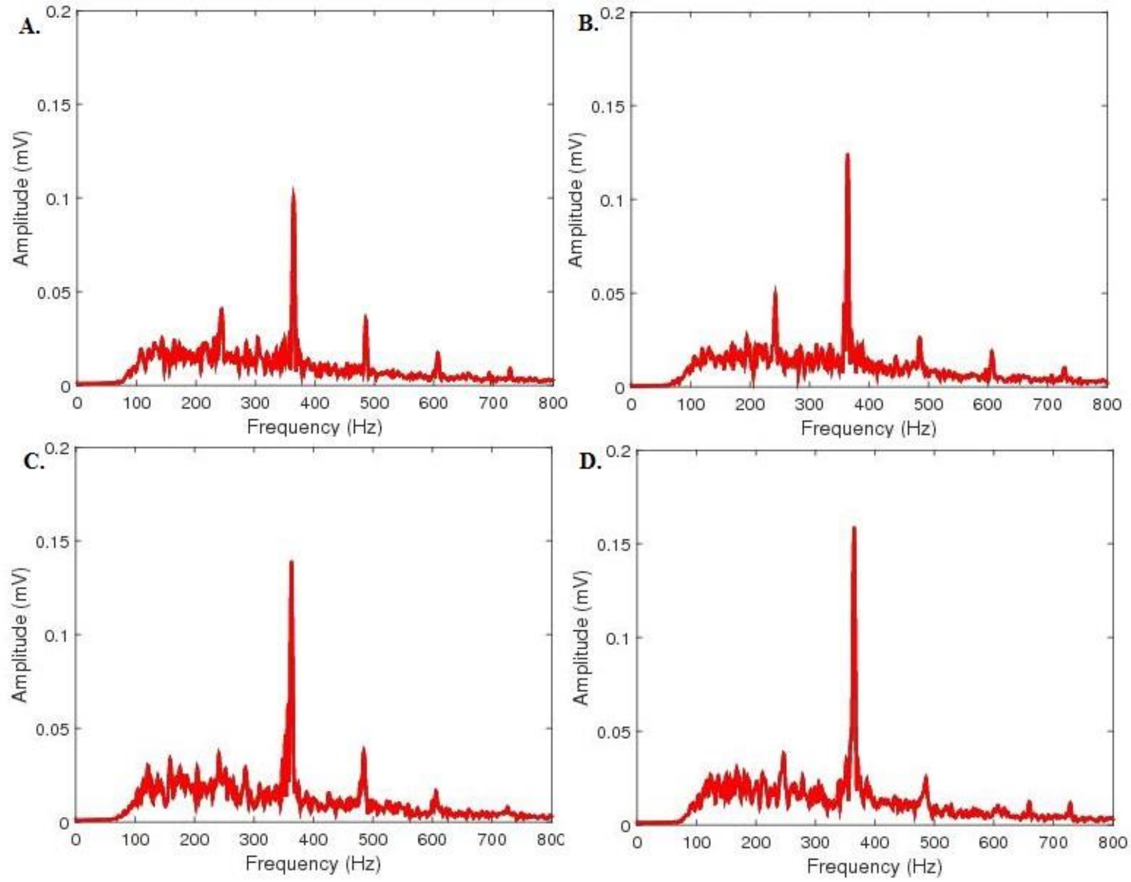

Figure 25. Grand-averaged FFR_{TFS} FFTs of six normal hearing subjects across four reverberation conditions. (Panel A: Clean Condition, Panel B: Mild Reverberation, Panel C: Moderate Reverberation, Panel D: Severe Reverberation).

Two key findings were observed for the FFR_{TFS} FFT data: (1) an unexpected increase in F1 amplitude as a function of reverberation severity; the peak corresponding to the F1 gradually increased in amplitude with increasing reverberation (panel A (0.109 mV), B (0.113 mV), C (0.134 mV), D (0.155 mV)); and (2) an apparent decrease in amplitude of the formant-related harmonics with increasing reverberation severity, especially in the moderate and severe reverberation conditions (panels C and D). The strongest F1 amplitude at 360 Hz was noted in the severe condition (panel D (0.155 mV)) and the smallest F1 amplitude was observed in the clean condition (panel A (0.093 mV)).

F1 Quantitative Spectral Analysis as a Function of SNR: Individual and Mean Values

The initial analysis of the F1 data was the same as that done for the F0 data. However, this analysis resulted in an unexpected finding which suggested that F1 amplitude values increased as the severity of the reverberation increased. For example, the absolute mean FFT amplitude of the F1 response in the clean condition (.093 mV) was smaller than the mean FFT amplitude of the F1 in the most severe condition (.155 mV). This finding is evident in both the individual and the mean data shown in Table 4 below. The variability seen in this F1 data was relatively similar across test conditions, and this cannot explain this finding.

Table 4

Individual F1 Component Absolute Amplitudes of Six Normal Hearing Participants for Varying Levels of Reverberation with Descriptive Measures.

	F1 Component				
Participants	Clean	Mild RT: 0.6	Moderate RT: 0.8	Severe RT: 1.1	Difference
S1	0.052 mV	0.064 mV	0.151 mV	0.143 mV	0.091 mV
S2	0.038 mV	0.048 mV	0.067 mV	0.084 mV	0.046 mV
S3	0.079 mV	0.091 mV	0.152 mV	0.123 mV	0.044 mV
S4	0.117 mV	0.166 mV	0.152 mV	0.205 mV	0.088 mV
S5	0.208 mV	0.244 mV	0.215 mV	0.255 mV	0.047 mV
S6	0.063 mV	0.065 mV	0.067 mV	0.119 mV	0.056 mV
M	**0.093 mV**	**0.113 mV**	**0.134 mV**	**0.155 mV**	**0.062 mV**
SD	**0.062 mV**	**0.076 mV**	**0.057 mV**	**0.063 mV**	**0.001 mV**

Note. Difference= difference between response amplitude of clean condition and severe reverberation condition.

One of the possible reasons underlying the unexpected finding described above could have been variations in the level of the noise floor in each of the test conditions. This may be a possibility because the FFT analysis is an absolute measurement which adds the desired signal (F1 encoding amplitude) in with the underlying noise floor. In order to investigate this possibility, noise floor values were analyzed across conditions.

The noise floor measurements indicated that there was indeed an increase in the level of the mean noise floor as the severity of the reverberation increased. For example, the mean noise floor measurements for the clean condition were smaller (.016 mV) than for the mild (.019 mV), moderate (.022 mV), and severe (.021 mV) reverberation conditions. In order to compensate for differences in the noise floor level across conditions and to normalize the data, the signal to noise ratio (SNR) was calculated for each test condition for all participants.

The SNR of the response controls for the noise floor by comparing the desired signal (which in this case would be F1 amplitude strength) with the present noise/unwanted signals. In doing so, a more accurate representation of the desired signal is obtained.

Analysis of the F1 SNR data continued to reveal inconsistent findings. The trend noted in both the absolute FFT (increased amplitude measurements as condition severity worsened) was observed for the mean F1 SNR values as well. The mean and standard deviation were calculated for F1 SNR values at each condition and are listed in Table 5 along with individual subject values.

Table 5

Individual F1 SNRs of Six Normal Hearing Participants for Varying Levels of Reverberation with Descriptive Measures.

	F1 Component				
Participant	Clean	Mild RT: 0.6	Moderate RT: 0.8	Severe RT: 1.1	Difference
S1	2.88 dB	3.13 dB	5.96 dB	6.12 dB	3.24 dB
S2	3.76 dB	3.48 dB	5.03 dB	4.97 dB	1.21 dB
S3	6.55 dB	6.93 dB	6.93 dB	8.08 dB	1.53 dB
S4	6.78 dB	6.14 dB	4.76 dB	8.60 dB	1.82 dB
S5	8.89 dB	10.50 dB	8.32 dB	9.45 dB	0.56 dB
S6	3.39 dB	3.82 dB	5.02 dB	5.34 dB	1.95 dB
M	**5.38 dB**	**5.67 dB**	**6.00 dB**	**7.09 dB**	**1.71 dB**
SD	**2.39 dB**	**2.82 dB**	**1.39 dB**	**1.86 dB**	**0.53 dB**

Note. Difference= difference between clean condition SNR and severe reverberation condition SNR.

As noted for the mean absolute F1 amplitude measurement, the mean SNR values increased as the reverberation condition worsened. For example, the mean SNR for the clean condition (5.38 dB) was less than the SNR for the severe condition (7.09 dB). The greatest increase in F1 SNR between sequential conditions appears to occur between the moderate (6.00 dB) and severe (7.09 dB) reverberation conditions. The smallest increase in F1 SNR seems to occur between the clean (5.38 dB) and mild (5.67 dB) reverberation conditions. The greatest variance in F1 SNR was observed for the mild reverberation condition (2.82 dB) whereas the smallest variance was observed for moderate reverberation condition (1.39 dB). The variability is considerably greater in the quiet and mild test conditions versus the moderate and severe test conditions. For example, the variance for the mild condition was twice as large as that observed for the moderate condition (2.82 dB/1.39 dB). Individual SNR values calculated at each test condition are in general agreement with mean values across conditions for all the subjects.

The trends discussed above can be visualized in Figure 27 below. This figure plots the mean F1 SNR obtained in each of the four conditions, with mean F1 SNR in dB on the y-axis and reverberation condition represented on the x-axis. Individual SNR values calculated at each test condition are in general agreement with mean values across conditions. Possible explanations for this pattern of results will be further discussed in the discussion section.

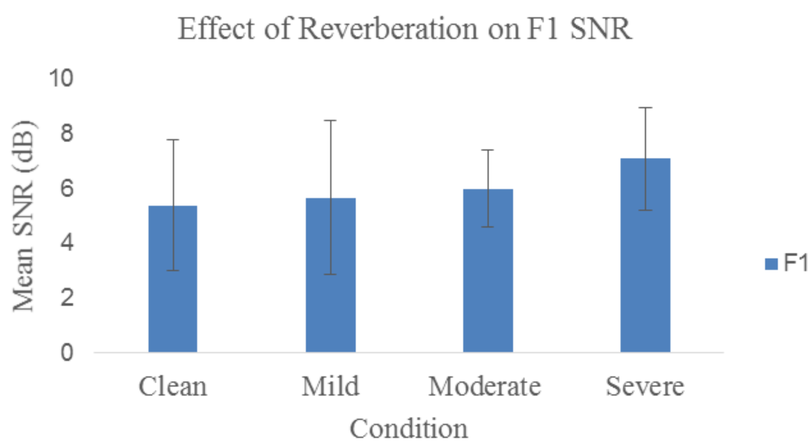

Figure 26. Grand-averaged mean F1 SNR across varying levels of reverberation. As reverberation severity increases, mean F1 SNR increases. Note the standard deviation bars.

Summary of Results

Overall, the results from the current experiment can be summarized as follows:

A. FFR_{ENV} Component Results:

  a. Temporal Waveform Composition

    i. Decreased temporal waveform amplitude with increasing levels of reverberation.

    ii. Decreased temporal resolution of waveforms with increasing levels of reverberation.

    iii. The highest degradation in temporal waveform amplitude and temporal resolution occurred in the most severe reverberant condition.

  b. Spectral Composition

    i. The presence of spectral energy at the F0.

    ii. The presence of spectral energy at the harmonics of the F0.

    iii. A substantial decrease in F0 amplitude in the severe reverberation condition.

  c. Quantitative Spectral Analysis

    i. A substantial decrease in F0 mean amplitude values with increasing reverberation.

B. FFR_{TFS} Component Results:

  a. Temporal Waveform Composition

    i. In general, only the severe reverberation condition showed notable change in waveform morphology.

    ii. An increase in temporal waveform amplitude as a function of reverberation severity.

  b. Spectral Composition

    i. Increased F1 amplitude as a function of reverberation severity.

    ii. Decreased amplitude of formant-related harmonics with increasing reverberation severity, especially in the moderate and severe reverberation conditions.

c. Quantitative Spectral Analysis

i. Increased mean SNR values as reverberation severity increased.

ii. Considerably greater variability in SNR in the quiet and mild test conditions versus the moderate and severe test conditions.

Collectively, all of these results relate to two general observations: (1) decreased brainstem encoding of F0 with increasing levels of reverberation, especially in the severe reverberation condition, and (2) an unexpected increase in brainstem encoding of F1 with increasing levels of reverberation.

CHAPTER 5: DISCUSSION

The purpose of this study was to determine the effects of reverberation on brainstem neural encoding by analyzing the FFR_{ENV} and FFR_{TFS} components in normal hearing individuals. Brainstem encoding ability was assessed using a /u/ speech stimulus with an F0 of 120 Hz and an F1 of 360 Hz. It was hypothesized that subcortical neural encoding of both F0 and F1 would decrease as a function of reverberation, with F1 encoding experiencing greater degradation. In examining the effects of reverberation on these FFR components, it was hoped that this information could shed some additional information regarding listening difficulties experienced by either normal hearing or hearing impaired individuals in degraded listening environments.

There were two primary findings in this study: (1) brainstem encoding of the F0 decreased with increasing levels of reverberation, especially in the severe reverberation condition; and (2) there was an unexpected increase in brainstem encoding of F1 with increasing levels of reverberation. However, in order to further discuss the results of the present study, a review of basic neural encoding provided below is necessary.

Resolved & Unresolved Harmonics

The observed effects of reverberation on F0 and F1 encoding can be linked to the underlying physiology of the cochlea (Sayles & Winter, 2008). This section will specifically review the neural encoding of complex signals via the auditory filters located on the basilar membrane and discuss how the working of these auditory filters relates to brainstem encoding of complex sounds such as the /u/ speech stimulus used in the present study. It is recommended that the reader use Figure 28 below as a guide to this topic. This figure illustrates a complex stimulus waveform at the top (A) with arrows below this stimulus showing the various harmonic components of this stimulus (B) passing through distinct areas of the auditory filters along the basilar membrane (C).

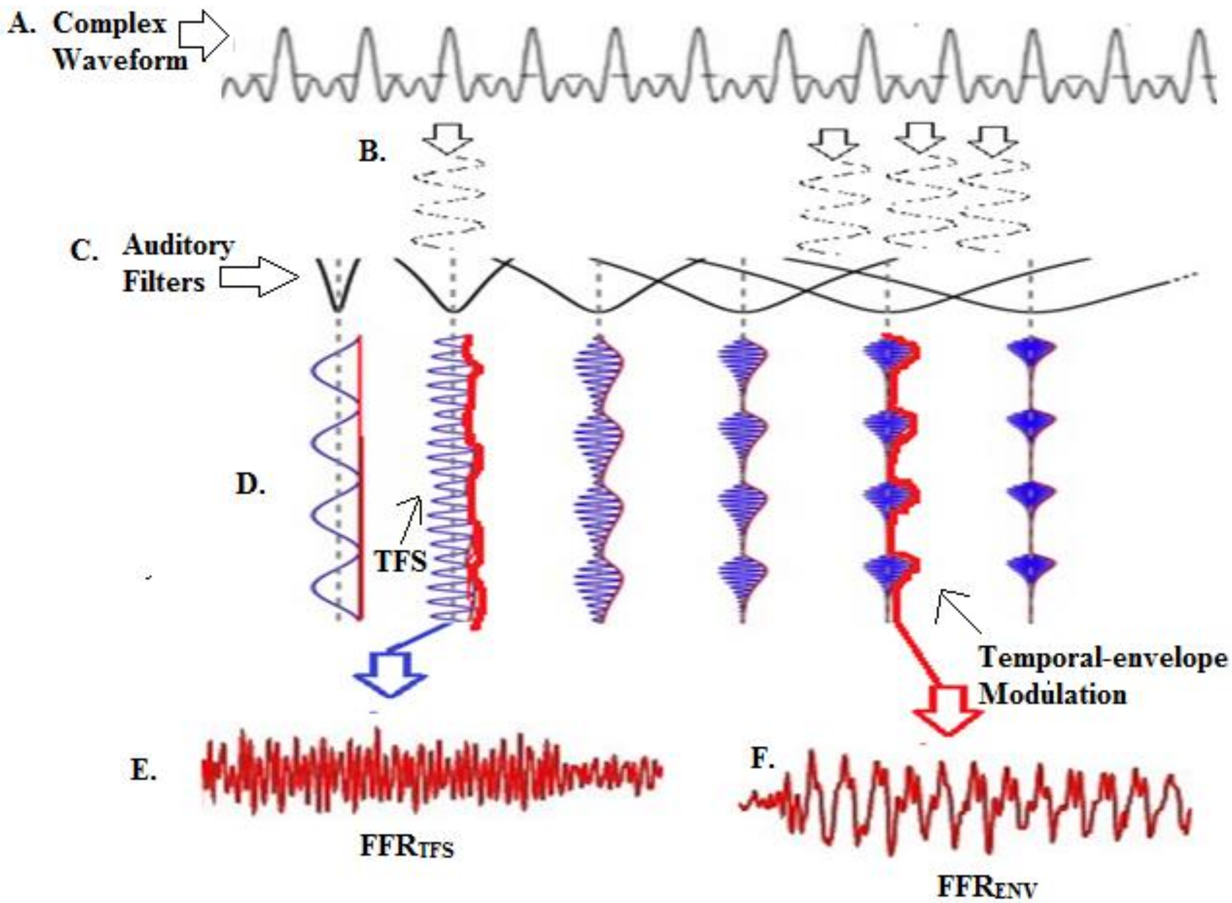

Figure 27. Basic illustration showing the difference between auditory filter output along the basilar membrane. Modified from "Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds" M. Sayles and I. Winter, 2008, *Neuron*, 58(5), 789-801.

The basilar membrane of the cochlea consists of multiple band pass auditory filters (C). These auditory filters are tonotopically organized depending on the location of the auditory filters along the basilar membrane, some are narrow and selective (as seen on left hand side of part C) or broad and less selective (as seen on right hand side of part

C). These auditory filters are each responsible for processing and isolating the many frequency components that make up complex sounds.

At the apical end of the cochlea, these auditory filters selectively process low frequency components of the signal through narrow band-pass filters. The resulting filter outputs are representative of sinusoidal waves known as “resolved” harmonics (Sayles & Winter, 2008). This can be visualized on the left side of part B in Figure 27 which shows a single component of the complex stimulus waveform entering the second auditory filter. The output of this auditory filter shows a clearly defined sine wave. However, at the basal end of the cochlea, these auditory filters overlap and process multiple higher frequency components at once. Therefore, the product of these auditory filters at this level of the cochlea are representative of complex waveforms consisting of “unresolved” waveform (Sayles & Winter, 2008). Such a complex waveform made of unresolved harmonics is pictured on the right side of part B and is the result of three component waveforms entering a broader auditory filter (the fifth filter from the left).

While each of these outputs contain TFS information, temporal-envelope modulation is only evident at the output of the higher frequency, wider filters which deal primarily with the unresolved harmonics. Each of these pieces of information are shown in Figure 27. The TFS (represented in blue) are the rapid oscillating complex waveforms seen at the output of each auditory filter, regardless of place along the basilar membrane. The temporal-envelope modulation, however, is only seen at the output of the higher frequencies, and is therefore specific to this higher frequency place along the basilar membrane. The envelope modulation at the output of the wide, high frequency filter dealing with unresolved harmonics is highlighted in red for the output at the fifth auditory filter.

The above information is fundamental to discussing the effects of reverberation on neural encoding because brainstem encoding occurs in response to both the resolved and unresolved outputs. The subsequent sections will now go into detail about the effect of reverberation on these outputs.

Effect of Reverberation on F0 Encoding

It has previously been reported that F0 encoding shows marked degradation with increasing reverberation (Sayles & Winter, 2008). The present study also found a substantial effect of reverberation on F0 encoding, especially in the severe reverberation condition. This decrease was evident in both the FFT amplitude measurements as well as in the qualitative temporal waveforms indices.

Understanding the auditory filter mechanism is crucial as it has been suggested that there are two ways in which the pitch of a complex stimulus can be determined by the auditory system (Bidelman & Krishnan, 2010). The first occurs by combining the neural phase-locking to the TFS for the resolved harmonics, as shown in part E of Figure 27. The second transpires by phase-locking to the temporal-envelope modulation which as discussed earlier, represents output of wider high frequency filters dealing with unresolved harmonics. An example of this phase-locking to the temporal-envelope modulation is shown in part F of Figure 27. Sayles and Winter (2008) have explained that of these two pitch extracting methods, the second is significantly degraded with increasing reverberation. This degradation is believed to be the result of randomization of the phase relationships that are present in unresolved harmonics (Bidelman & Krishnan, 2010). Therefore, the ability of the auditory system in determining pitch is severely compromised when a complex sound containing only unresolved harmonics is presented in reverberation.

This piece of information is critical to our discussion as it has been suggested that the brainstem F0 encoding is not place specific and is reliant on the envelope of higher frequency unresolved output for F0 encoding (Greenberg, Marsh, Brown, & Smith, 1987). This was initially suggested by Smith, Marsh, Greenberg and Brown (1978), who recorded the FFR in humans using both a pure tone and complex tone with a missing fundamental. This study used masking noise to mask the spectral information of the fundamental frequency (365 Hz) of a pure tone and a complex stimulus. When masking noise was introduced into the pure tone stimulus, the FFR was significantly reduced. On the other hand, when masking noise was introduced at the place of the F0 of the complex stimulus, a robust FFR was produced. These researchers concluded that because the FFR was still obtained to the complex tone stimulus with a missing fundamental, that the underling pitch of the FFR is mediated by a place other than the one that approximates to the F0 of the stimulus.

This concept has been maintained in a recent unpublished study by Ananthakrishnan (2013), where F0 encoding in response to an intact /u/ stimulus was found to be stronger in individuals with normal hearing sensitivity versus individuals with sloping high frequency hearing loss. In an attempt to understand this difference, normal hearing individuals in this study were also presented a /u/ stimulus with missing high frequency information, to replicate the audiometric configuration of hearing impaired listeners. Interestingly, normal hearing responses to this modified stimulus were similar to those obtained in the individuals with high frequency hearing loss to an intact stimulus. This finding suggests that access to high frequency information is a major contributor to F0 encoding, and when this information is limited, the result is a decrease in F0 encoding ability.

Collectively, the above information emphasizes three key points: (1) unresolved harmonics are more susceptible to reverberation effects than lower frequency resolved harmonics; (2) F0 encoding in the FFR is dependent on unresolved harmonic information; (3) and F0 encoding is affected by reverberation because it relies on unresolved harmonic information. Therefore, the decrease in F0 encoding with increasing reverberation that was observed in the present study is in agreement with these points.

The mean FFT amplitude measurements that were obtained are generally consistent with the spectral findings of Bidelman and Krishnan (2010), who found robust encoding at the F0 for the clean condition, with minimal/slight decrease at the mild and moderate reverberation conditions, and truly significant reduction at the F0 for the severe reverberation. The present study showed decreased periodicity in grand-averaged temporal waveforms as a function of reverberation. These results are consistent with Bidelman and Krishnan (2010), who observed the greatest F0 encoding in a clean condition and the lowest F0 encoding ability in a severe reverberation condition.

Lastly, the decrease in F0 encoding that was observed in the present study can also be potentially related to a spectral phenomenon that was discussed in the literature review. Nabelek et al. (1982) suggested that overlap masking, a phenomenon that occurs in vowel sounds, causes a significant number of errors in speech identification tasks in normal hearing individuals. Since the present study used a vowel stimulus, it is possible that overlap masking effects resulted in noted smearing of the temporal waveform response, especially in the severe reverberation condition where periodicity and waveform amplitude were visibly reduced. This smearing could also explain why there was also increased difficulty in distinguishing the F0 peak from the peaks of the harmonics in the FFT for the severe reverberation condition.

Effect of Reverberation on F1 Encoding

Sayles and Winter (2008) noted in their animal model, that neuronal units tuned to lower frequencies show significantly less degradation in phase-locking ability in reverberation than units tuned to higher frequencies, which have been shown to change even in mild levels of reverberation. Whereas F0 encoding occurs at the output of the filters with higher center frequencies, F1 encoding occurs in a more place specific manner. Specifically, brainstem F1 encoding is mediated by either resolved or unresolved harmonic information, depending on the actual value of the F1 frequency. The frequency of the first formant determines whether it is processed through a narrow low center frequency filter, resulting in a resolved harmonic, or a wide high center frequency filter, resulting in an unresolved harmonic. The F1 of the stimulus in the present study is a relatively lower frequency harmonic that passes through narrow filters and which emerges as a resolved harmonic. Therefore, brainstem F1 encoding in this study occurred in response to this resolved harmonic, in a place specific manner.

As previously mentioned, Ananthakrishnan (2013) recorded the FFR recorded using a /u/ stimulus in both normal hearing and hearing impaired individuals. Specifically, this study found that when both an intact stimulus and a low pass filtered version of the stimulus were presented to the normal hearing group, there were no notable differences in F1 encoding between each type of stimulus. On the other hand, the hearing impaired group showed reduction of F1 encoding to the intact stimulus. These findings further support the notion that F1 encoding occurs in response to the resolved output of filters with lower center frequencies. Based on this and on the points made in the earlier portions of the discussion section, there are also three key points that can be inferred about F1 encoding. These are the following: (1) resolved harmonics are not susceptible to the effects of reverberation; (2) F1 encoding is mediated by resolved harmonics; (3) and F1 encoding is not affected by reverberation because it relies on resolved harmonic information. Results from the present study indicated that there was no decline in F1 encoding with increasing reverberation. These results are consistent with the key findings related to F1 encoding that have been discussed.

Although there was no decrease in F1 encoding in the present study, there was however, an increase in F1 encoding across reverberation conditions. This finding is not in agreement with what has been suggested in the literature (Bidelman & Krishnan, 2010), who found that the formant-related harmonics were significantly reduced across reverberation conditions. This effect of reverberation on F1 encoding does not correspond to the trend demonstrated in the F1 data of the present study. Although the discussion thus far has focused primarily on the role of cochlear auditory filters as a reasoning for the effects of reverberation on F0 and F1 encoding, this explanation still does not provide a clear reason as to why there was an improvement in F1 encoding as reverberation severity increased. It should, however, be noted that individual and mean F1 encoding data showed a greater degree of variability than F0 encoding data. This variability is noteworthy because the present study only tested six participants. If more participants had been tested and shown increased F1 encoding, these results could be analyzed with less caution. However, due to the small sample size, the validity of the F1 encoding data is poor. Therefore, it could very well be that the observed F1 trend occurred as a result of variability effects and not reverberation.

Limitations of the Study

There were several limitations in the present study. This study recorded the FFR in six normal hearing adults aged 23-25 years. This is a relatively small cohort with a small range of ages. Therefore, findings from the current study are limited and cannot be generalized to other populations (i.e., hearing impaired individuals, elderly individuals, etc…). Due to testing only a small number of participants (n=6), the actual statistical analysis for this study was restricted to descriptive statistics. Therefore, no significant differences or effects could be calculated for each condition. Previous studies that have used the FFR to study brainstem encoding in degraded listening conditions have included a larger number of participants, typically greater than 10-12 participants and have been able to find significant effects (Anderson, Skoe, Chandrasekaran, & Kraus, 2010; Bidelman & Krishnan, 2010; Cunningham et al., 2001). In addition, only one speech stimulus (/u/ vowel) was used to obtain the FFR to varying levels of reverberation. Because speech encompasses many sounds (consonants, diphthongs, and vowels), it would have been of value to utilize a greater number speech stimuli in order to obtain responses that give a more accurate representation of every day speech.

Clinical Relevance and Future Directions

While there have been many studies that have looked at the effect of reverberation on speech perception, the vast majority have been behavioral studies (Duquesnoy & Plomp, 1980; Nábělek, 1988; Nábělek & Dagenais, 1986; Nábělek, Ovchinnikov, Czyzewski, & Crowley, 1996; Nábělek & Robinson, 1982). To date there is limited objective research on the effect of reverberation on subcortical neural encoding of speech in humans (Bidelman & Krishnan, 2010). The present study looked to add to this limited objective research. Results from this study can serve as a stepping stone in explaining why hearing aid users experience increased listening difficulties in adverse listening conditions. Further FFR research is needed in the hearing impaired population in order to compare their results to those obtained in the normal hearing population. Once these differences are further understood, hearing aid manufacturers can focus on developing spectral technology for patients who require greater listening benefits in adverse listening conditions.

Conclusions

Results from this study suggests that reverberation, regardless of severity causes a decrease in F0 neural encoding ability. However, reverberation effects are less substantial for F1 encoding. These differences in neural encoding ability between F0 and F1 components seem to be related to differences in the encoding of these two frequency components at the physiological level of cochlea.

Appendix A

INFORMED CONSENT FORM

I, ___________________________, agree to participant 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 will be listening to speech under both background noise and reverberation in varying conditions, and the intensity of the sounds will neither be painful nor intense.

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 form 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; as standard electrophysiologic techniques will be employed and the procedure is noninvasive and similar to that of being examined via EEG in a medical practice. The intensity level of the stimuli will be calibrated prior to beginning to any data collection. The intensity used is equivalent to normal to loud conversational level and therefore will not cause discomfort or harm. If I find that the intensity of the stimulus is uncomfortable, I understand that I can inform the investigator and testing will be discontinued. 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 am a student in the Audiology program at Towson University, withdrawing from or refusing to participate in the study will have no effect on my educational standing in the department.

If I have any questions or problems that arise in connection with my participation in this study, I should contact the primary investigator and Au.D. doctoral candidate, Randi Cropper at rcropp1@students.towson.edu / 240-298-2746, Dr. Saradha Ananthakrishnan, the thesis chair at sananthakrishnan@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

FFR PROTOCOL SHEET

SUBJECT # ________ INITIALS: ________

Before subject arrives:

- Turn on laptop for IHS system, and USB box under laptop
- Open IHS program
- Select **SmartEP**
- Select **Patient** from the toolbar
    - Choose **New** or **Open**
    - **OK** when complete
- Select **Stimulus** from the top toolbar
    - Select **Modality → Auditory- Advanced Research Module →ABR**
    - Choose **No** when asked about using default settings
- Select **Setup Advanced** on the bottom left of the screen. This will open a window.
    - Turn channel ON and browse to select stimulus
        - For **Reverb** you only need one right channel on:
            - Choose between u_clean, u_mild, u_mod, and u_severe
        - For **BGN** turn both right channels ON:
            - Select u_clean for the first right channel
            - Choose between 0 SNR, +5 SNR, and -5 SNR
    - Alternating polarities
    - Data points: **4096**
    - Stimulus rate: **3.13/s**
    - Sampling rate: **75**
    - Artifact rejection end: **307550**
    - When adjustments complete, select **OK**
- Set page: Left plot end time and then enter **300 ms**
- Double click sweeps: **2000**
- Open **EEG & Amplifier** settings from bottom right toolbar
- Check each electrode channel you plan to use, and ensure the designation is sent to **ON** and not left or right
    - Remember: Channel **A** is for linked mastoids
    - Remember: Channel **B** is for C7
- Ensure the earphones are emitting a stimulus
- Take out subject randomization sheet and mark the initials of the subject in the next available subject slot

When the subject arrives:

Day 1 start time: ______________ **Day 2 start time: ______________**

- Give a basic explanation of the experiment, goals, procedures

- Give subject consent form, have him/her sign it
- Wash hands
- Otoscopy
- Prepping
  - Scrubbing using NuPrep gel and electrode paste/tape to secure electrodes according to color scheme
    - Ground (Low Forehead) → **Green**
    - Fpz (High Forehead) → **Brown**
    - A_2 (Right mastoid) → **Red**
    - A_1 (Left mastoid) → **Grey**
    - C7 (Neck) → **Peach**
      - Linked mastoids in channel A
      - C7 in channel B
- Impedence check
- Re-instruct the patient (include time of test: ~1.5 hours)
- Insert earphone; ask if the sound is comfortable
- **Acquire**

<u>Data Acquisition</u>

Random order of stimuli but include:

- BGN Session: Clean, +5 SNR, 0 SNR, -5 SNR
- Reverb Session: Clean, Mild RT, Mod RT, Severe RT
  - Remember to monitor power spectrum
    - Process → Power spectrum
- Use randomization sheet to label next to list below

	Start Time	*Stop Time*	*Artifacts*	*# Runs*	*Comments*
Clean					
+5 dB SNR					
0 dB SNR					
-5 dB SNR					

Background Noise Session: DAY 1/ DAY 2 (circle one)

Reverberation Session: DAY 1/ DAY 2 (circle one)

	Start Time	*Stop Time*	*Artifacts*	*# Runs*	*Comments*
Clean					
Mild RT					
Mod RT					
Severe RT					

General notes:

After data collection:

- Remove inserts
- Remove electrodes/ clean up participant
- Close out

Day 1 end time: _______________ **Day 2 end time: ______________**

APPENDIX C

IRB APPROVAL LETTER

TOWSON
UNIVERSITY

APPROVAL NUMBER: 15-A054

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

From: Institutional Review Board for the Proctection 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. Ananthakrishnan; P. Korczak
File

TOWSON UNIVERSITY

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.

(for)

Scot McNary, Member
Towson University Institutional Review Board

References

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1), 35-47.

Akhoun I, Gallego S, Moulin A, et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology, 119*, 922–933.

Ananthakrishnan, Saradha, "Neural Encoding of Complex Signals in the Healthy and Impaired Auditory Systems" (2013). *Open Access Dissertations*. Paper 198. http://docs.lib.purdue.edu/open_access_dissertations/198

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(14), 4922–6. doi:10.1523/JNEUROSCI.0107-10.2010

Arai, T., Nakata, Y., Hodoshima, N., & Kurisu, K. (2007). Decreasing speaking-rate with steady-state suppression to improve speech intelligibility in reverberant environments. *Acoustical Science and Technology*, *28*(4), 282-285.

Arthur, R. M., Pfeiffer, R. R., & Suga, N. (1971). Properties of two-tone inhibition in primary auditory neurons. *The Journal of Physiology*, *212*(3), 593-609.

Atcherson, S.R. & Stoody, T.M. (2012). *Auditory Electrophysiology: A Clinical Guide*. New York: Thieme Medical Publishers.

Banai K, Hornickel J, Skoe E, et al. (2009). *Reading and Subcortical Auditory Function*. Cereb Cortex, *19*, 2699–2707. [PubMed: 19293398]

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

Boothroyd, A. (2004). *Room Acoustics and Speech Perception*. In Seminars in Hearing. *25*, 155-166. Thieme Medical Publishers Inc.

Burkard, R. F., Eggermont, J. J., & Don, M. (Eds.). (2007). *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins.

Chandrasekaran, B., & Kraus, N. (2010). The scalp recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology, 47*(2), 236-246.

Chiappa, K. H. (1997). *Evoked Potentials in Clinical Medicine*. Lippincott-Raven. Retrieved from http://books.google.com/books?id=h5mOjnTimT8C

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hearing Research, 264*(1), 48-55.

Crandell, C. C., & Smaldino, J. J. (2000). Classroom acoustics for children with normal hearing and with hearing impairment. *Language, Speech, and Hearing Services in Schools, 31*(4), 362-370.

Cunningham J, Nicol T, Zecker SG, et al. (2001) Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology, 112*, 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.

Daube, J. R., & Rubin, D. I. *Clinical Neurophysiology*. 2009.

Davis, H., & Hirsh, S. K. (1976). The audiometric utility of brain stem responses to low-frequency sounds. *International Journal of Audiology, 15*(3), 181-195.

Davis, H., & Hirsh, S. K. (1979). A slow brain stem response for low-frequency audiometry. *International Journal of Audiology*, *18*(6), 445-461.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain Event-Related Potentials in Man* (pp. 349-441). New York: Academic Press.

Duquesnoy, A.J., & Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis. *The Journal of the Acoustical Society of America, 68*(2), 537–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7419810

Emanuel, D. C., & Letowski, T. (2009). *Hearing Science*. Wolters Kluwer Health/Lippincott Williams and Wilkins.

Febo, D. M. (2003). *Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing* (Doctoral dissertation, University of South Florida).

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., Amaya, E. M., de Rivera, J. M. D., Donan, N. M., Duong, M. T., Hsu, J. N., & Tsang, L. P. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport*, *15*(13), 2057-2060.

Galbraith, G. C., Arbagey, P. W., Branski, R. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, *6*(17), 2363-2367.

Galbraith, G. C., & Arroyo, C. (1993). Selective attention and brainstem frequency-following responses. *Biological Psychology, 37*(1), 3-22.

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., & Mullen Jr, T. A. (1998). Brain stem frequency - following response to dichotic vowels during attention. *Neuroreport, 9*(8), 1889-1893.

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, 19*(3), 203-214.

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.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport, 14*(5), 735-738.

Gardi, J., Merzenich, M., & McKean, C. (1979). Origins of the scalp-recorded frequency-following response in the cat. *International Journal of Audiology, 18*(5), 353-380.

Glaser, E. M., Suter, C. M., Dasheiff, R., & Goldberg, A. (1976). The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalography and Clinical Neurophysiology, 40*(1), 25-32.

Gnanateja, G., Ranjan, R., & Sandeep, M. (2012). Physiological bases of the encoding of speech evoked frequency following responses. *Journal of the All India Institute of Speech & Hearing, 31*.

Goblick Jr, T. J., & Pfeiffer, R. R. (1969). Time ‐ domain measurements of cochlear nonlinearities using combination click stimuli. *The Journal of the Acoustical Society of America, 46*(4B), 924-938.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology, 12*(6), 767-782.

Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research, 25*(2), 91-114.

Hall, J. W. (2007). *New handbook of auditory evoked responses* (pp. 123-136). Pearson.

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), 437-443.

Hodoshima, N., & Arai, T. (2007). Effect of talker variability on speech perception by elderly people in reverberation. In *Proc. International Symposium on Auditory and Audiological Research*.

Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility. *Journal of Speech, Language, and Hearing Research, 33*(4), 726-735.

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 and Hearing, 32*(6), 699-707.

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C. D., & Chung, H. K. (2010). Early maturation of frequency following responses to voice pitch in infants with normal hearing. *Perceptual and Motor Skills, 111*(3), 765-784.

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology, 92*(4), 311-324.

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *The Journal of Neuroscience, 28*(15), 4000-4007.

Katz, J. (2009). *Handbook of Clinical Audiology*. Lippincott Williams & Wilkins.

Killion, M. C. (1997). Hearing aids: past, present, future: moving toward normal conversations in noise. *British Journal of Audiology, 31*(3), 141-148.

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters, 319*(2), 111-115.

Koppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. *The Journal of Neuroscience, 17*(9), 3312-3321.

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*(1), 192-201.

Krishnan, A. (2007). Frequency-following response. *Auditory evoked potentials: Basic principles and clinical application*, 313-335.

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), 1-12.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161-168.

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*(1), 21-27.

Lipscomb, David M. *Noise and Audiology*. Baltimore: University Park Press, 1978. Print.

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

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. (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, 9*(4), 399-406.

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology, 35*(6), 665-667.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*(40), 15894-15898.

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.

Nábělek, A. K. (1988). Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. *The Journal of the Acoustical Society of America, 84*(2), 476–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3170940

Nábělek, A. K., & Dagenais, P. A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners. *The Journal of the Acoustical Society of America, 80*(3), 741–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3760327

Nábělek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *The Journal of the Acoustical Society of America, 86*, 1259. Retrieved from http://link.aip.org/link/?JASMAN/86/1259/1

Nábělek, A. K., Ovchinnikov, A. C., Yewski, Z., & Crowley, J. (1996). Cues for perception of synthetic and natural diphthongs in either noise or reverberation.

The Journal of the Acoustical Society of America, 99(3), 1742–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8819863

Nábělek, A. K., & Robinson, P. K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *The Journal of the Acoustical Society of America, 71*(5), 1242–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7085996

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience, 29*(45), 14100-14107.

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. USA: Allyn and Bacon.

Picton T.W. *Auditory evoked potentials*. In Daly DD & Pedley TA (eds.) *Current Practice of Clinical Electroencephalography*, 2nd edn. New York: Raven Press, 1990, pp. 625–678.

Raphael, L. J., Borden, G. J., and Harris, K. S (2008). *Speech Science Primer: Physiology, Acoustics and Perception of Speech*, 5th Edition.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology, 115*(9), 2021-2030.

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research, 156*(1), 95-103.

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. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing, 31*(3), 302.

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*(5), 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.

Snyder, R. L., & Schreiner, C. E. (1985). Forward masking of the auditory nerve neurophonic (ANN) and the frequency following response (FFR). *Hearing Research, 20*(1), 45-62.

Stevens, S. S., and Hallowell Davis. *Hearing, Its Psychology and Physiology*. New York, N.Y.: American Institute of Physics for the Acoustical Society of America, 1983. Print.

Stillman, R. D., Crow, G., & Moushegian, G. (1978). Components of the frequency-following potential in man. *Electroencephalography and Clinical Neurophysiology, 44*(4), 438-446.

Wever, E., & Bray, C. (1930). Auditory nerve impulses. *Science, 71(*1834), 215. doi: 10.1126/science.71.1834.215-a

Wile, D., & Balaban, E. (2007). An auditory neural correlate suggests a mechanism underlying holistic pitch perception. *PloS One*, *2*(4), e369.

Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42-52.

CURRICULUM VITA

NAME: Donald Guillen

[redacted]

PROGRAM OF STUDY: Audiology

DEGREE AND DATE TO BE CONFERRED: Doctor of Audiology (Au.D.), May 2016

Secondary Education: Bethesda-Chevy Chase High School, Bethesda, MD, June 2008

Collegiate institutions attended	Dates	Degrees	Date of Degree
Loyola University Maryland	8/2008-5/2012	B.A.	5/2012
Towson University	8/2012-5/2016	Au.D.	5/2016

Professional positions held:

ENTAA Care- Columbia, MD
Audiology Student Clinician, January 2015-May 2015

Johns Hopkins Bayview Medical Center- Baltimore, MD
Audiology Student Clinician, September 2014-December 2014

Richard H. Israel, Ph.D., Audiology Services- Silver Spring, MD
Audiology Student Clinician, June 2014-July 2014

The Hearing and Speech Agency- Baltimore, MD
Audiology Student Clinician, January 2014-May 2014

Towson University Speech, Language and Hearing Center- Towson, MD
Audiology Student Clinician, September 2012-December 2013