The Perception and Neural Representation of Individual Harmonics in a Vowel Sound: A Behavioral and Auditory Brainstem Evoked Response Study

by

Jessica Andrus

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2011

DALHOUSIE UNIVERSITY

SCHOOL OF HUMAN COMMUNICATION DISORDERS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “The Perception and Neural Representation of Individual Harmonics in a Vowel Sound: A Behavioral and Auditory Brainstem Evoked Response Study” by Jessica Andrus in partial fulfillment of the requirements for the degree of Master of Science.

Dated: December 16, 2011

Supervisor: ___________________________________

Readers: ___________________________________

____

____

DALHOUSIE UNIVERSITY

DATE: 16 December 2011

AUTHOR: Jessica Andrus

TITLE: The Perception and Neural Representation of Individual Harmonics in a Vowel Sound: A Behavioral and Auditory Brainstem Evoked Response Study

DEPARTMENT OR SCHOOL: School of Human Communication Disorders

DEGREE: MSc CONVOCATION: May YEAR: 2012

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions. I understand that my thesis will be electronically available to the public.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than the brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

____
Signature of Author

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Vowel perception primarily depends on the overall shape of the speech spectrum, which is imposed by the positions of the primary speech articulators. Voiced vowels also have a harmonic fine structure due to the activity of the vocal folds, and these harmonics give rise to synchronized activity in the brainstem. This synchronous firing may be useful for speech perception in noise and speaker discrimination, although it is unknown if the synchronized neural response to the harmonic increases perceptual audibility of the harmonic. The focus of the current study was to examine the relationship between the audibility of harmonics and the brainstem response to harmonics. The individual harmonics were found to be encoded in the brainstem, determined using brainstem frequency-following response recording, and the individual harmonics were audible to the individual, as determined using the pulsation threshold technique; however there was minimal relationship between the frequency-following response and perception of harmonics.

LIST OF ABBREVATIONS USED

FFR	Frequency –Following Response
ASSR	Auditory Steady State Response
dB	Decibel
dB HL	Decibel hearing level
Hz	Hertz
2AFC	Two Alternative Forced Choice
FFT	Fast Fourier Transform
DP	Distortion Product
H	Harmonic number
F	Formant number
DPOAE	Distortion product otoacoustic emission
DP-FFR	Distortion product frequency following response

ACKNOWLEDGEMENTS

I’d like to acknowledge the time and effort to this project committed by my supervisors, Dr. Steve Aiken and Dr. Michael Kiefte. This research would not have happened without their knowledge and wisdom of the topic, and their ability and willingness to share that knowledge with me. Thank you for the continued patience and support with my constant questions, frequent emails, and panicked meetings.

Thank you to Dr. Jian Wang, who served on the thesis committee and offered valuable advice along the way. I’m also grateful for the feedback and advice that will be provided by my external examiner.

Thank you to those who encouraged me to continue on with this research when I wasn’t quite sure I had it within me. I appreciate your honest thoughts and advice.

Thanks to my family for being there for me every step of the way, and feigning interest in my thesis research despite it reading like a pile of gibberish to you. Thank you for the faith and support; when I doubted my abilities, your faith in my abilities reminded me that I can do it.

Thank you to my friends. Those of you in research, we shared moments of excitement, anticipation, triumph, and frustration. For all my friends, thank you for listening to my frustrations when I needed to talk. You were always there for a cup of coffee or a glass of wine, depending on the day.

Finally, thank you to my husband for being there for me; being there with a cup of coffee when I needed a boost, a gentle push when I stalled, and a hug and smile as we went about our busy days. I wouldn’t be who I am without you, so thank you for supporting me through this process.

CHAPTER 1 INTRODUCTION

One of the most valued uses of our hearing is for communication, and spoken language depends on a person's ability to detect, discriminate, and perceive the individual sounds of the language. Although the past decade has shown an increase in awareness of hearing impairments and advances in technology for the hearing impaired, persons with hearing loss still face severe difficulties in understanding speech in daily environments that present non-ideal listening conditions. Understanding the roles of the individual components of the speech signal for speech perception is essential to the rehabilitation and treatment of the hearing impaired.

One of the methods used to understand the roles of the components of the speech signal in perception is to study the neural processing of speech. An example of a frequently used speech evoked response is the mismatch negativity, which occurs roughly 150-250 ms after the onset of a deviant sound imbedded in a series of standard sounds. The MMN is used as a measure of preattentive speech processing. However, there is a risk associated with studying the neural processing of speech as an indicator of speech perception because the relationship between neural responses to speech and the process of speech perception is not always clear. The current research focused on the relationship between the audibility of individual harmonics in a vowel stimulus and the synchronized brainstem activity to the same harmonics, as measured by the frequency-following response.

1.1 SPEECH ACOUSTICS

The sounds that compose spoken language are separated into vowels and consonants. While consonants are articulated with a complete or partial closure of the vocal tract, a vowel is produced with an open vocal tract. The identification of a vowel depends primarily on the location of spectral peaks corresponding to resonances of the vocal tract, which are influenced by the relative positions of the speech articulators (the tongue, the jaw, and the lips). An individual resonance peak of the sound spectrum for a particular vowel is called a formant, and a particular vowel is characterized by the frequencies of these formants. In addition to the formant frequencies, all voiced speech has energy at integer multiples of the fundamental frequency (determined by glottal pulse rate) called harmonics; these are produced by the activity of the vocal folds in the larynx. The amplitudes of individual harmonics are determined by the formant location, because formant resonances affect the level of the speech harmonics occurring in the same frequency range.

The role of formants in the perception of speech sounds has been explored extensively. Peterson and Barney (1952) recorded vowel utterances from 76 speakers, including men, women and children. They presented the recordings to a group of listeners, and asked them to identify the vowel sound. Scatterplots of first and second formant frequency showed that the vowel sounds formed clusters on the graph; this suggests that the first and second formant frequencies help to define the phonetic value of the vowel. Liberman and colleagues (1954) found that the first and the second formant frequencies provide sufficient information for the listener to correctly identify vowels, while the remaining formants did not increase the success with vowel identification. These studies suggest that the first and second formant values of vowel sounds contribute to the overall perceptual identity of the vowel, and are sufficient to lead to accurate vowel perception. The crude shape formed by the formants, which reflects the positioning of the articulators, is thought to be a very important feature for vowel perception.

The basic model of vowel perception founded on the research outlined above by Peterson and Barney (1952) and Liberman and colleagues (1954), suggests that F1 and F2 frequency alone are sufficient for vowel identification.   More recent research suggests that vowel identification is not that basic; we are unable to accurately predict a listener’s vowel identification based on formant frequencies alone (Rosner and Pickering, 1994). These findings highlight a need for alternative models of vowel perception (see Rosner and Pickering, 1994 for a review of the research). Some authors suggest that a more broad analysis of the vowel spectrum occurs during vowel perception and that an integration of the spectral peaks over a broad frequency range is responsible for vowel identification (Blandon and Lindblom, 1981; Ito *et al.,* 2001). Aaltonen (1985) found that by manipulating the amplitude of F2 while maintaining all other parameters of the stimulus constant, the listener’s vowel identification changed from the original vowel stimulus to a new identity. By reducing the amplitude of formant peaks, the vowel identity changes; this could be due to a release from simultaneous masking of higher formants by lower frequency formants, and/or by a change in the degree of spectral contrast of a formant (spectral contrast is the difference between peaks in the spectrum).

Kiefte and colleagues (2010) examined the influence of formant amplitude on vowel perception and the relative roles of simultaneous masking and spectral contrast by using both full spectrum stimuli or incomplete spectrum stimuli (only harmonics at or adjacent to formant frequencies were present). They manipulated the amplitude of F2 in the vowel /u/, and asked the participants to identify the vowel. When F2 was detectable, the stimulus was identified as /u/; when it was inaudible, the stimulus was identified as /i/. They found that simultaneous masking and local spectral contrast likely both play some role in vowel perception. The importance of spectral contrast was demonstrated in the difference in the effect of F2 amplitude between full spectrum stimuli and incomplete spectrum stimuli. With the incomplete stimuli, there was maximal spectral contrast as the harmonics between the formant harmonics were removed, so any observed changes were due only to masking. Masking also plays a role in vowel perception because listeners misidentified some vowels in the incomplete spectrum condition, despite that F2 was always present. Kiefte and colleagues (2010) also indirectly demonstrated that a listener only needs the harmonics around the formants to correctly identify a vowel stimulus, since their participants accurately identified vowels with an incomplete spectrum. Kakusho (1971) concluded that only 2-3 harmonics near formant frequencies were sufficient for 100% accurate vowel identification. Both of these findings suggest that harmonics between formant peaks are not important in vowel perception.

The auditory system is characterized by very high temporal and spectral resolution. If all that is necessary for vowel perception is the comparison between the first few formant frequencies, it is unnecessary to have the perceptual ability to precisely resolve the harmonic frequencies in the speech stimulus. While it seems that the overall shape of the speech spectrum (formed by the formants) is important for speech perception, the role of the fine harmonic detail and our ability to perceive it is not well understood. Smith and colleagues (2002) created auditory chimeras that have the envelope (i.e., formant shape) of one speech sample but the fine details (i.e., harmonics) of another speech sample. In these studies, it was found that the envelope of the speech – speech auditory chimera was most important for speech identification (perception of words), but the harmonics were most important for pitch details and sound localization. For example, if the speech envelope was from sample 1 and the harmonic details were from sample 2, the listeners invariably perceived the words from sample 1 and not sample 2. In coherence with previous studies, this study suggests that the formants provide the necessary information for basic speech understanding. However, this study suggests that while the harmonic fine structure does not directly contribute to speech comprehension, it provides additional information that can facilitate speech understanding in a variety of contexts. While speech can be perceived in the absence of the harmonics (e.g. in whispered speech) by deciphering the formants, vowel perception may use the harmonics as additional information if they are present.

The ability to separate sounds in noise may depend on the ability of the auditory system to resolve the important fine details in speech (i.e. the harmonic structure). A single speaker’s formants are carried by a distinct harmonic trajectory, which can be used to track a speaker in auditory space (Bregman, 1993). Also, harmonic resolution is necessary for speaker discrimination, and unless the speech from simultaneous speakers can be separated, speech understanding will be poor. In general, the harmonics in speech likely play an important role in auditory scene analysis (Bregman, 1993), where sounds are separated, discriminated and localized to a source.

The importance of harmonics in speech perception is particularly relevant to persons with a sensorineural hearing impairment. A characteristic of sensorineural hearing loss is outer hair cell damage, which reduces the tuning of the auditory system due to the loss of the outer hair cell active mechanism (Gelfand, 2004). Sensorineural hearing loss is characterized by reduced sensitivity and reduced frequency selectivity as a result of broadened auditory filters; this may reduce the resolution of the individual harmonics of speech. If individual harmonic resolution is found to improve the clarity of speech in noise, it is not unreasonable to suggest that the decrease in selectivity and sensitivity would cause deficits in speech intelligibility for a person with a hearing impairment.

1.2 ELECTROPHYSIOLOGICAL CORRELATES OF HARMONIC PERCEPTION

In order to make use of any additional information provided by the harmonics of a speech stimulus, the auditory system must be able to detect and encode the harmonics. An objective method to determining if the speech harmonics are being represented is to examine the neural processing of the speech harmonics. Speech sounds trigger both a transient and an on-going neural response in the brainstem and the cortex, and analyzing the response of the brainstem to harmonics could verify that the harmonic information is being encoded by the auditory system and is therefore available for use by the auditory cortex during speech perception.

The neurons of the central auditory system respond to onsets, offsets, and changes of speech stimuli, and this gives rise to electrical potentials that can be measured at the scalp; these can be used to verify the encoding of the speech sounds in the neural system but they do not resemble the actual characteristics of the stimulus, only the timing of the events (these are called ‘transient responses’). Brainstem neurons are also capable of generating steady-state responses as a result of the neuronal firing rate synchrony to the stimulus components, and this response directly reflects the components of the speech stimulus. When a population of auditory neurons fires in synchrony, a potential change at the scalp can be measured. In general, the transient brainstem response can be used to indicate that the stimulus component was represented by the auditory system, while the steady state response can provide more information about the actual representation of the component in the auditory system.

Steady state responses can occur to the modulation rate of a stimulus with constant amplitude and frequency content over time (e.g., the 40 Hz auditory steady state response), or to the actual frequency of a tone rather than its modulation (e.g., the frequency-following response). A typical example of an FFR is the cochlear microphonic, which mimics the waveform of the stimulus almost perfectly and with little latency. The brainstem neurons also generate an FFR, with a significantly longer latency of about 15 ms due to the onset delay of the FFR. The FFR decreases in amplitude with increasing frequency and becomes difficult to recognize above 1500 Hz (Moushegian, Rupert & Stillman, 1973). This is termed the low pass filter of the brainstem, and it is the result of the limits of phase locking in a population of auditory neurons.

Speech is a naturally modulated stimulus in the time and spectral domain, with an overall waveform structure with a few larger peaks (formants), and smaller waves embedded within the formant peaks (harmonics). As mentioned previously, the harmonics of voiced speech represent the distortion produced by the saw-tooth like movement of the vocal folds. The vocal folds open gradually, and then close rapidly, which combine to create an overall amplitude fluctuation. Not only are the harmonics present in the acoustic stimulus, they also create an overall amplitude modulation by interacting in the cochlea and auditory neurons during analysis.

The brainstem following response that is synchronized to the temporal envelope of the speech stimulus is generally called an envelope-following response (Aiken and Picton, 2006), or an auditory steady-state response (Dimitrijevic *et al.,* 2004), but it has also been described as a frequency-following response (Krishnan *et al.,* 2004). More commonly, the frequency-following response describes responses to frequency components in the stimulus (e.g., Aiken & Picton, 2006). Aiken and Picton (2008) distinguished between the two responses by calling the frequency-following response to changes in the temporal envelope (fluctuations in stimulus energy over time) and the FFR to harmonics as the envelope-FFR and the spectral-FFR respectively. It has also been suggested that the FFR may reflect formant encoding as well.

Plyler and Ananthanarayan (2001) used the frequency-following response to describe a response to the formants in a speech stimulus. However, given that formants do not have an inherent temporal structure, the responses were likely related to the harmonics in the vicinity of the formant. Speech stimuli can be used to evoke frequency-following responses in the brainstem to the fundamental frequency of the stimulus or to the harmonic details of the speech stimulus. Krishnan (2002) showed that the envelope FFR is insensitive to the polarity of the stimulus, while the spectral FFR is sensitive to the polarity of the sound stimulus; this provides a method of distinguishing between the frequency-following responses by recording responses to stimuli presented in opposite polarities and averaging the difference between the responses (Huis in't Veld et al., 1977; Yamada et al., 1977). There are multiple techniques that have been used to confirm that the brainstem provides an on-going response to the fundamental frequency of a natural speech stimulus. Aiken and Picton (2006) recorded responses to the fundamental frequencies of natural vowel stimuli with either steady or changing fundamental frequencies. They used a sine-cosine Fourier analyzer to measure the energy in the response to the fundamental as it changed over time (called the frequency trajectory). Using the Fourier analyzer, a significant response was elicited in all subjects. Krishnan *et al*. (2004) used an autocorrelation algorithm to obtain speech FFRs to Mandarin Chinese stimuli with dynamic fundamental frequencies, and Dajani *et al.* (2002*)* recorded speech FFRs using a filter-based algorithm that mimicked cochlear physiology. However, although measuring responses to the fundamental frequency is quick and reliable, it does not provide detailed information about the audibility of different frequencies in speech; all energy in voiced speech is amplitude-modulated at the fundamental frequency. In order to get more frequency specific information, the frequency-following response to individual harmonics must be recorded.

Krishnan (2002) recorded the frequency-following response to low-frequency synthetic vowels and used a fast Fourier transform to provide an estimate of the response to each harmonic in the stimulus. Half of the stimuli were polarity-inverted, and the final response was obtained by subtracting the response to the original stimulus from the response to the polarity-inverted stimulus. Harmonics close to formant peaks and other low frequency harmonics elicited significant responses, while no response was recorded to the fundamental frequency. This subtractive approach teases apart the different results for the envelope FFR and the spectral FFR.

There are different averaging procedures for the alternating polarity frequency following recordings that can be used to look at the different components of the response. Responses elicited by the envelope (envelope-FFRs) were distinguished from responses related to the spectral details (spectral-FFR) by adding/subtracting the responses recorded to the vowel stimuli in opposite polarities. In the nomenclature, the first sign indicates if the responses are added (+) or subtracted (-), and the second sign indicates the polarity of the second set of responses (- indicates an inverted polarity). The first set of response is always the original stimulus. The average response to stimuli presented in the original polarity (++ average) is composed of both the spectral and the envelope FFR. If the FFRs recorded to opposite polarities are added (+-), the response is small and composed predominantly of the envelope FFR, distortion products, and electrical noise. When FFRs to opposite polarities are subtracted (--), the result is the best indication of the spectral components (harmonics) in the stimulus because it removes any rectification-related distortion (including the envelope of the stimulus). The spectral-FFR is clearest in the -- average, however it is present in the ++ and +- average but confounded by the envelope FFR and distortion products. The envelope FFR is present in the ++ and the +- average, but absent in the -- average; subtracting opposite polarity FFRs eliminates the envelope-FFR because the inverted polarity has minimal effect on the envelope.

The harmonics and fundamental frequency of natural stimuli change across time, so the response to the harmonics of natural speech is not optimally analyzed using the same fast Fourier transform method traditionally used to determine the response to the fundamental frequencies of synthetic, static speech. Instead of the fast Fourier transform, the Fourier analyzer can be used to calculate the energy of dynamic signals. Aiken and Picton (2008) recorded the response to the natural vowel /ɒ/ and calculated the spectrum using both the fast Fourier transform and the Fourier analyzer to compare the two methods. They found that the harmonic amplitudes were underestimated by the fast Fourier transform, suggesting that the Fourier analyzer is a more accurate method for determining the audibility of harmonics that do not have a constant frequency.

In Aiken and Picton (2008), the spectral FFR showed significant responses to the harmonics closest to the formants, with the peak response corresponding to the harmonics surrounding the first formant. These findings suggest that the auditory system is able to represent individual harmonics, and that this temporal representation may be available for further processing, but this does not imply that the auditory cortex will complete further processing of the stimulus using these temporal cues. Even if an individual harmonic is perceived, and there is brainstem representation of the harmonic, the temporal representation may not play a role in the processing; it could be a rate-place code that is responsible for the processing. Exploring the relationship between the temporal code and vowel processing was a focus of the present study.

One consideration during electrophysiological studies are distortion products, as it can be difficult to determine if the measured response is a true neural response or distortion created during the processing of a signal. A distortion product is an intermodulation between two frequencies in a nonlinear system, such as the auditory system. The intermodulation of two frequency components creates additional sound components at various frequencies. Non-linear distortion products are created in the speech signal prior to inner hair cell transduction, so if a distortion product is measured, it will be reflected in the neural encoding of the stimulus. The movement of outer hair cells (Brownell, 1990) and the resultant motion of the basilar membrane can cause distortion products in response to a stimulus. Also, the mechanics of the hair cell stereocilia can create distortion products (Liberman et al., 2004). These distortion products can be measured with minimal latency from the cochlea as otoacoustic emissions (Miller et al., 1997), or directly from the central auditory system using electrophysiologic recordings (Purcell et al., 2007). Miller and colleagues (1997) detected neural responses to speech harmonic distortion products in the phase locking of the auditory nerve, while Krishnan (2002) recorded distortion products in the human FFR.

Krishnan (2009) examined the effects of off-frequency masking of F1 on F2 by varying either the amplitude of F1, or by increasing the F2 frequency and recording the FFR. In the original stimulus with the natural F1 amplitude, the waveform was composed of mostly low frequency energy, reflecting both the peak at F1 and a DP-FFR. As the level of F1 decreased, the FFR at F1 and the DP-FFR also decreased; however, the energy at F2 increased with decreasing values of F1. If the levels of F1 and F2 were held constant but F2 was increased in frequency, an increase in F2 amplitude was noted for increases in F2 frequency. Both of these experiments showed a release from masking on F2 by F1. The DP-FFR also exhibited predictable behavior as F2 was increased in frequency; F2-F1 DP-FFR shifted higher when F2 increased in frequency, and 2F1-F2 decreased in frequency as the frequency of F2 increased. Krishnan (2009) also recorded frequency following responses to predictable distortion products in response to a digitally synthesized two tone vowel approximation.

If DP-FFRs are present in the response to a vowel, they could have a variety of effects on the speech FFR. Harmonics in the acoustic stimulus and related distortions can overlap at the harmonic frequencies, thus making it difficult to resolve which components are responses to the stimulus and which are by-products of the nonlinearity of the system.

Since cochlear distortion products are not related to the half-wave rectification during the signal processing, they are not removed by subtracting the alternating polarity responses, so they should be present in the spectral-FFR. The amount of response due to distortion products must be examined when using the FFR to determine the response to individual harmonics; this can be examined by looking at the effect of removing an individual harmonic from the stimulus. If the response to the harmonic continues to occur in a stimulus that is missing the harmonic, this suggests that the response is predominantly due to distortion products.

1.3 PSYCHOACOUSTICAL CORRELATES OF HARMONIC PERCEPTION

In order to understand the importance of individual harmonics to speech perception, it is necessary to determine the perceptual audibility of the individual harmonics in complex signals by the individual. The finding that the brainstem responds to harmonics only indicates that the lower portion of the central auditory system encodes the fine spectral details, and not necessarily that the auditory cortex uses the information during perception. There have been attempts to determine the audibility of individual harmonics using psychoacoustic masking techniques, but all are time-consuming, difficult, and not ideally suited to the audiology clinic. These methods have been based on the assumption that the threshold of the signal will be affected by the presence of a masker. Vogten (1974) demonstrated that low-frequency sounds can mask high frequency target stimuli, when the masking stimulus has sufficient amplitude. Within a vowel, a low frequency formant may be capable of masking the next highest formant and the harmonics in between. The masking techniques include forward masking (Moore and Glasberg, 1983), pulsation threshold (Houtgast, 1974), and simultaneous masking (Moore and Glasberg, 1983).

In simultaneous masking (Moore and Glasberg, 1983), the listener is presented with the target stimulus and the masking stimulus concurrently. The masker is fixed in intensity, and the target amplitude is adjusted to a level where it is just audible in the presence of the masker; this intensity level is termed the masking level of that target stimulus. This paradigm presents many confounding factors, including masker-target interactions such as combination tones and beats that may lower the estimate of the actual signal threshold.

Forward masking is a form of non-simultaneous masking where the masker is presented first, and immediately after the masker is stopped, the target stimulus is presented (Moore and Glasberg, 1983). Again, the level of the masker is held constant and the level of the target stimulus is adjusted to the masking threshold.

A common masking technique to study vowel representation in the auditory system is the pulsation threshold technique. In this paradigm, the masker (the vowel) is pulsed on and off, alternating with the target stimulus (a pure tone) during the masker silent periods. Typically, the level of the masker is held constant. The theory behind this technique is based on Thurlow's (1959) discovery that a weak tone alternated with a strong tone will sound like it is pulsing, except when the amplitude of the weak tone is low enough that it is masked by the stronger tone. At this low level of the weak tone, the presentation sounds continuous, as if the weak tone was extending through the strong tone. This effect was termed "auditory induction" by Warren *et al*. (1972), and it occurs when the intensity difference between two alternating sounds are such that the weaker sound is completely masked by the stronger sound. This results in the perception of a continuous tone. Therefore, at low stimulus levels in the pulsation threshold technique, the signal is perceived as a continuous tone in the presence of a pulsing masker. If the signal is raised higher beyond a fixed threshold, the participant will no longer hear a continuous tone, but one which alternates with the masker. The sound level at which the perception of the target signal changes from continuous to pulsating is called the pulsation threshold. Houtgast (1973) found that the masking functions obtained from this paradigm correspond to the physiological estimates of the frequency selectivity of the peripheral auditory system. Thus, by using the pulsation masking patterns, the auditory representation of the speech spectrum can be inferred behaviourally.

Macintosh and Kiefte (2005) used the pulsation threshold technique to determine the audibility of individual harmonics within a vowel. One harmonic at a time was chosen to be the target signal, and the remaining vowel spectrum was employed as the masker. Two variations of the technique, the self-adjustment method (Houtgast, 1974) and the two alternative forced choice method (Levit, 1971), were compared. They found that in general, harmonics between formant peaks were not audible, while harmonics adjacent to formant peaks were audible. It is also important to note that although the same trends were observed in both variations of the method, the two-alternative forced-choice method led to the most accurate estimations of the thresholds.

1.4 OVERVIEW OF CURRENT STUDY

This study looked at the extent to which a normal listener can hear individual harmonics in vowel sounds, and evaluated the contribution (if any) of the temporal encoding in the brainstem to the processing of the harmonic. The study determined whether the ability to hear a specific harmonic relates to the degree of neural synchronization to that harmonic. It may be possible to obtain a quick electrophysiologic measure that would provide information about the audibility of individual harmonics, which could be adapted to a short clinical test that would make it possible to assess the individual audibility of harmonics in a short time in the clinic with difficult-to-test clients. This study tested this idea by recording pulsation thresholds and steady-state (frequency-following) brainstem responses to vowels.

The two-alternative forced-choice method used by Macintosh and Kiefte (2005) was used in the present study to verify the audibility of the individual harmonics. The listener was presented with a pulsating stimulus containing a target stimulus (the harmonic of interest), and the masking stimulus (all remaining harmonics of the vowel spectrum). They were forced to choose which of the two presentations was continuous, and one presentation was truly continuous while the other was pulsating. If they chose the correct stimulus twice in a row, the pulsating target stimulus intensity level was decreased by 2 dB (masker is fixed at 76 dB SPL). If they incorrectly chose once, the pulsating signal was increased by 1dB. The listener incorrectly chose the pulsating signal when they could no longer hear the difference between the truly continuous signal and the pulsating one, which was called the threshold. The masking threshold of a particular harmonic indicates the audibility of that individual harmonic; a low masking threshold indicates that the level of the harmonic in the original signal was above the masking threshold and thus clearly audible. If the harmonic had a high masking threshold, it suggested that the harmonic was not audible in the vowel stimulus. By using these masking paradigms, we were able to determine if the listener could perceptually hear that individual harmonic in the vowel stimulus.

The audibility of the harmonic in the vowel spectrum, as determined by the psychoacoustic masking procedure, was compared to the neural representation of the harmonic in an electrophysiology investigation. The method used by the present study to investigate the neural activity in response to an individual harmonic is similar to the method used by Aiken and Picton (2008). The stimulus was the vowel recording, presented with each of the first 13 harmonics removed (at any given time, the stimulus had a single harmonic absent). The brainstem FFR to harmonics is difficult to interpret because the response at any harmonic frequency could be a result of energy at that frequency or a cochlear distortion product. Due to the potential contamination of the response at a given harmonic by distortion products, we removed each of the harmonics and derived the response at a given harmonic both with it present and absent. This recognized the concern that any response recorded to an individual harmonic may be the result of distortion products. Electroencephalographic recordings were made while the participants relaxed in a chair and listened to the stimuli. By adding and subtracting responses recorded in opposite polarities, we were able to differentiate responses to envelope and harmonic information in the speech (Aiken and Picton, 2008). The frequency component representation in the response recorded to a vowel (same recording as used for the behavioural task) was determined using a fast Fourier transform. The neural firing patterns in response to changes in the harmonic content of the stimulus were examined to determine the ability of the harmonic to elicit a synchronized neural response.

Each participant's individual performance on the behavioural task was compared to the response to the corresponding harmonic in their brainstem recording. If performance in the behavioural task indicated that the harmonic was audible in the vowel stimulus for a given participant (low masking threshold), this study examined whether there was a corresponding change in the brainstem response when that harmonic was removed from the stimulus. Theoretically, if the brainstem response was synchronized to the harmonic and the person indicated that they could detect the presence of the harmonic, the harmonics could be providing information that increased speech intelligibility.

1.5 POTENTIAL IMPLICATIONS OF THE STUDY

The outcomes of the study have implications for both diagnostic testing and rehabilitation techniques. If the neurophysiologic results strongly correspond with the behavioural results, than this is a test of harmonic audibility that could be useful in a clinical setting in difficult testing situations. Clinicians use general speech testing to determine the audibility of speech, however there are clients who cannot reliably provide responses to speech testing. The FFR approach could be valuable in estimating the audibility of speech in infants or hard to test clients. Finally, the present research could have implications for the frequency resolution of hearing aids and the frequency processing strategies of cochlear implants. If we understand the role that harmonics play in speech discrimination, we can improve our techniques for hearing rehabilitation to increase the functional outcomes of people with hearing impairments.

CHAPTER 2 METHODS

2.1 PARTICIPANTS

Complete data was obtained from fifteen normal hearing adult listeners all with audiometric thresholds better than 20 dB HL at octave frequencies from 0.25 to 8.0 kHz (i.e. no significant hearing loss). These participants were unpaid volunteers, and they were recruited internally at Dalhousie University by the author.

All procedures used in this study received ethical approval by the Dalhousie Research Board prior to testing (Protocol #2010-2197). This experiment posed no emotional or physical risk to the participants.

2.2 BRAINSTEM RECORDING PROCEDURE

2.2.1 Stimuli

The formant frequencies of a live male voice recording of the vowel /ɒ/ (f0=114 Hz), whose first formant was 790 Hz and second formant was 1195 Hz, was used to create a synthetic vowel stimulus using MATLAB implementation of the Klatt 80 speech synthesizer (Klatt, 1979). Vowel spectrum can be found in Figure 1 of the appendix, and the harmonic and formant frequencies are in Table 1 of the appendix. The vowel /ɒ/ was chosen because it has a low frequency second formant, maximizing the chances of recording a second formant response.

The stimulus was manipulated so in a given trial, a single harmonic was missing from the stimulus. The first thirteen harmonics were removed (past the second formant) in ascending and descending order for each participant and these stimuli were concatenated into a sweep stimulus. Each sweep contained 28 separate stimuli, each of which was presented for 75 ms, with a 75 ms offset-to-offset interstimulus interval. The first stimulus was the complete /ɒ/ vowel, followed by versions of the /ɒ/ missing a single harmonic. The harmonics were deleted in order from the first (the fundamental) to the thirteenth, and then in reverse order from the thirteenth to the first. The full stimulus was presented a second time at the end of the sweep. Sweeps were presented in alternating polarity. All stimuli were presented at a level of 76 dBA, as measured in a 2 cc coupler.

During electrophysiological testing, the stimuli were presented through Labview and an M-series data acquisition device, at a sample rate of 32 kHz. The digital stimuli were presented monaurally with an Ear-Tone 3A insert (300 Ohms impedance) into the participant's right ear. All stimuli were presented at a level of 76 dBA as determined in a 2 cc coupler.

2.2.2 Recordings

The electroencephalographic recordings were obtained while participants were relaxed or sleeping in the sound booth. Responses were recorded between gold electrodes at the vertex and the mid-posterior neck, and a ground electrode was behind the left ear. Responses were digitized at 10,000 Hz using the same M-series device and custom LabVIEW software. The responses were amplified 10,000 times, and filtered between 30 Hz and 3000 Hz by a Grass LP-511 biopotential amplifier before digitization.

2.2.3 Procedure

While resting or sleeping, the listener was presented with recording blocks and their FFR was recorded to the stimulus sweeps. In each recording block, there were 214 stimulus sweeps (107 in each polarity), for a recording duration of 14 minutes and 59 seconds. Each block was presented four times to the listener, for a recording time of roughly 1 hour. The participant was monitored throughout recording for any changes in artifact level or changes in responses.

2.2.4 Data Analysis

Due to our use of a synthetic stimulus with no variation over time, we used a Fast Fourier Transform to determine the energy in the frequency following response to each individual harmonic, from H1 to H13 for both polarities of the stimulus. For each participant, an average response to each polarity of the stimulus was calculated for each of the 4 blocks. The recording period was 150 ms, during which the stimulus occurred in the first 75 ms and the last 75 ms was the silent period. In order to obtain the frequency following response to the stimulus, the difference in the response to the two polarities was averaged; by calculating the difference average, the data consists of the portion of the response that changes when the polarity of the stimulus changes (frequency following response). The frequency following response was calculated using the difference average and averaged across all four blocks. After the frequency following response was obtained for the 150 ms recording window, the portion of the response occurring to the stimulus was extracted. The FFR has a slight onset delay, so the response occurring 15 ms after the stimulus onset until the offset of the stimulus (75 ms) was extracted, providing 60 ms of data.

In order to disentangle the response components related to the stimulus harmonics, the frequency associated with the amplitude envelope, and any distortion products, both a +- average and a - - average was used. As outlined in the introduction, the - - average eliminates the envelope FFR and represents the spectral-FFR, any distortion products, and any noise. The +- average eliminates the CM and the majority of the spectral-FFR, while preserving the envelope-FFR.

2.3 PSYCHOACOUSTIC PROCEDURE

2.3.1 Stimuli

The identical vowel stimulus as used in electrophysiological recordings was used for the psychoacoustic task. The digital stimuli were presented using the Tucker-Davis Technologies digital signal processor, and presented monaurally with an Ear-Tone 3A insert into the participant's right ear.

2.3.2 Two Alternative Forced Choice Method

One harmonic at a time was removed for the target signal, and the first thirteen harmonics were examined. The test harmonic served as the target, while the remainder of the vowel spectrum acted as the masker. The signal and the masker were delivered in alternating fashion at an interval of 75 ms on, and 75 ms off . The intensity of the masker was fixed at 76 dB A and the intensity of the signal was varied with the test.

In the two-alternative forced choice procedure, the participant was presented with two trials of the masker plus the signal in a pair. The trials began at a signal presentation level of 84 dB, while the intensity of the masker was constant at 76 dB A. In one trial, the signal was continuous and the masker was pulsed 3 times in 75 ms intervals. In the other presentation, the signal and the masker alternated in 75 ms intervals. The order of presentation of the continuous and alternating trials was random. The participant was instructed to choose the trial in which the signal was presented continuously. The masking threshold was determined by using a common "two-down one-up" procedure (Levit, 1971). If the participant chose correctly twice in a row, the intensity of the signal was decreased by 2 dB. If the participant chose incorrectly a single time (indicated that the pulsed signal is the continuous presentation), the intensity of the signal was increased by 1 dB. If the listener chose either the pulsed signal or the continuous signal with equal probability, it suggested that the listener could not tell the difference between the true continuous signal and the pulsed presentation, which indicated that they were at their masking threshold. The masking threshold is where the vowel masks the signal.

2.3.3. Data Analysis

The masking threshold for a single harmonic was calculated using the average of the last twelve reversals. Each participant has 13 masking thresholds, one for each harmonic tested. A threshold of 0 dB represents a harmonic that is just audible in the stimulus. A negative threshold indicates a harmonic is audible in the stimulus; it implies the level of the test harmonic in the original vowel stimulus was loud, and had to be decreased in order to be masked to determine a masking threshold. A positive threshold indicates the harmonic is not audible in the vowel; it suggests that the test harmonic must be increased to overcome the masking by the vowel and determine the masking threshold.

CHAPTER 3 RESULTS

3.1 BEHAVIORAL DATA ANALYSIS

The group average threshold for each harmonic was calculated and plotted based on relative perceptual salience. A masking level of 0 dB represents a harmonic that is just audible in the stimulus, a negative threshold indicates a harmonic that is audible in the stimulus, and a positive threshold represents a harmonic that is not audible in the original vowel stimulus. The masking levels for each harmonic are plotted in relative dB in Figure 2. This demonstrates that the most perceptually salient harmonic was H7, which is also the center harmonic of F1. H8 and H9 were inaudible, suggesting that they were being masked by upward spread of masking by the harmonics in F1. The low frequency harmonics (H1-H6) were also audible, reinforcing the role of the low frequency harmonics in encoding vocal pitch. H10 and H11 are harmonics contained by F2, and while they were audible in the original stimulus, their perceptual salience is reduced compared to H7. This could be due to masking effects of F1 on the frequencies of F2.

3.2 ELECTROPHYSIOLOGICAL DATA ANALYSIS

The Fast Fourier Transform is an algorithm that was used to calculate the response amplitudes at the harmonic frequencies in the stimulus. These are the raw amplitudes at each harmonic for the full stimulus; these values include the FFR to the harmonic, any distortion products, and noise. These were averaged across participants, and the amplitude for each harmonic is shown in Figure 3. The largest FFR responses are found at H7 and H8 (the harmonic frequencies contained by F1) and H10, a frequency of F2. There is less of a response at the higher frequencies, which is expected due to the low pass function of the brainstem.

Responses were recorded to the stimulus with individual harmonics removed, and the grand average response to the 13 stimuli with a missing harmonic (missing H1-13) can be found in Figure 4. It appears as though the response to a single harmonic is entirely related to the energy at that harmonic in the stimulus; once that harmonic is removed from the stimulus the response is greatly reduced. An alternate method to examine this portion of the data is the grand average response spectra (Figure 5). The black line is the response to the full stimulus, and the red line is the response to the stimulus with a single harmonic missing. The missing harmonic begins with 1 on the top, down to 13 on the bottom of the graph. In general, the energy at each harmonic in the response appears to be entirely dependent on the energy at the harmonic in the stimulus.

The envelope response is obtained by averaging the sum of the two alternating polarity recordings. The alternating polarity grand average across harmonics (Figure 6) shows a response at the fundamental frequency and the lower harmonics. This energy is reduced slightly when H1 is removed, and also when the harmonics near the formant peaks are removed.

3.3 COMPARATIVE ANALYSIS

The grand average FFR amplitude was compared to the grand average pulsation threshold for the same harmonic. The correlation between the grand average amplitude and grand average pulsation threshold across harmonics was minimal at r= -0.195. Based on the method of defining an audible harmonic (with a negative pulsation threshold), it is expected that any correlation suggesting a relationship between audibility and neural representation would be negative. As the perceptual salience of an individual harmonic increases, the pulsation threshold will decrease to a negative number; if high perceptual salience predicted a large FFR response, a negative correlation would reflect this relationship. However, there is no evidence of such a relationship with a correlation of

-0.195. The grand average amplitude versus threshold across harmonics is plotted in Figure 7.

The performance on individual harmonics was also examined across listeners. The 15 individual FFR amplitudes for a given harmonic were compared to the corresponding pulsation threshold for the same harmonic, yielding 13 correlations representing the relationship at each harmonic. Again, correlations were very low, with the highest correlation between behavioral threshold and FFR amplitude occurring at H4 (-0.428) and H7 (-0.414). This suggests little to no relationship between perception of the harmonic and the degree of neural representation in the brainstem.

The performance of individual participants was also examined. Figures 8-9 depict the harmonic amplitudes for the participant with the highest average FFR (best waveform response) and their corresponding pulsation thresholds for each harmonic. While the highest perceptual salience value matches the highest FFR amplitude at H7, there is no other predictable relationship at the other harmonics. The participant with the lowest average FFR amplitude is presented in Figures 10-11; with this participant, although H7 is the most perceptually salient, the highest FFR response is to H4.

CHAPTER 4 DISCUSSION

The purpose of this study was to evaluate the extent to which a normal listener can hear individual harmonics in vowel sounds, in an attempt to discover if the resolution of individual harmonics could contribute to the assignment of sound identity. The study attempted to determine whether the ability to hear a specific harmonic relates to the degree of neural synchronization to that harmonic.

4.1. INTERPRETATION OF BEHAVIORAL DATA

The pulsation threshold results suggest that the most perceptually salient harmonic was the center harmonic of the first formant of the vowel (H7). This is not a surprising result, as it has been shown that harmonics at or adjacent to formant frequencies are sufficient for vowel perception (Kiefte *et al.,* 2010; Kaksuho *et al.,* 1971), and it could be assumed that in order to contribute to vowel perception, it must have a strong representation during perception of the vowel. H7 is also the most intense harmonic in the acoustic stimulus. However, H8 and H9 were inaudible despite being fairly intense in the stimulus, likely due to simultaneous masking of the higher harmonics by F1. The louder, lower frequency harmonics contained by F1 masks the softer, higher frequency harmonic (Vogten, 1974). The tail of the traveling wave for F1 in the cochlea would pass by the area of representation for the higher frequency harmonics on the organ of Corti, and within a certain distance of the F1 frequency, the cells for the higher harmonics would still be occupied by the tail frequencies of F1. This would reduce the responsiveness of the hair cells and corresponding spiral ganglion neurons to the higher frequency harmonics, as they are being occupied (masked) by the high amplitude, lower frequency F1 (often referred to as upward spread of masking).

Simultaneous masking has been one of the factors suggested to account for observed effects of formant amplitude on vowel perception (Nearey and Levitt, 1974). This study does find a reduction in audibility of higher frequency harmonics adjacent to a formant, likely due to a masking effect of the first formant on these higher frequency harmonics. There is also evidence that this masking effect continues to mask the harmonic frequencies contained by F2 (H10 and H11), as their perceptual salience is reduced compared to H7 despite being center frequencies of the second formant. Krishnan (2009) documented an electrophysiological example of the masking of F1 on F2; when the amplitude of F1 was reduced, the FFR in response to F2 increased, suggesting a release from masking of F2.

The harmonics below the first formant (H1-H6) also had fairly strong perceptual representation, reinforcing the importance of lower frequency harmonics in determining vocal pitch. Vocal pitch is determined by the fundamental frequency of the voice (H1), therefore it is expected that H1 is very audible during the perceptual analysis of a speech stimulus. It appears as though there is slight downward spread of masking of the formant on H5 and H6; while H5 and H6 have more energy in the original spectrum than H3 and H4, they are less perceptually salient than H3/H4. While the upward spread of masking of low frequency sounds masking high frequency sounds is more typically observed, higher frequency sounds can mask lower frequency sounds if they are close enough in frequency.

In the original vowel spectrum, H12 and H13 are very low intensity, and the pulsation thresholds reflected this as H12 and H13 were inaudible. Kiefte and colleagues (2010) documented that harmonics between formants are not important to vowel identification by using incomplete spectrum stimuli in a vowel identification task. The incomplete spectrum only had harmonics at or adjacent to formant frequencies present in the stimulus, and this had no effect on the accuracy of vowel identification when the performance was compared to the full spectrum stimulus condition. Thus, it is not surprising that the H12 and H13 (harmonics between the second and third formants in the vowel stimulus /ɒ/) were not audible in the pulsation threshold task because of their intensity in the original stimulus and their relative unimportance in vowel perception.

4.2 INTERPRETATION OF ELECTROPHYSIOLOGY DATA

The most intense harmonics in the acoustic stimulus were H7 and H10, and these are the center harmonics of F1 and F2 respectively. The largest FFR was recorded to H7 and H8, followed by H10. The harmonic frequencies contained in the first formant (H6-H8) all had significant representation in the brainstem response. There is less of a response at the higher frequencies, which is expected due to the low pass function of the brainstem as outlined in the introduction. In general, FFR response amplitudes across harmonics mimic the levels in the original vowel spectrum, with the most intense harmonics in the stimulus showing the greatest response in the FFR.

When responses were recorded with single harmonics missing, in order to determine the effect of removing a harmonic from the spectrum, it was found that the energy at each harmonic in the response appears to be almost entirely dependent on the energy at the harmonic in the stimulus. This was an unexpected finding, as it was expected that the missing harmonic would be filled in to a certain degree by cochlear or neural distortion products. In the present study, when a harmonic was removed, there was no evidence of distortion products “filling in” the space for any participant. Any FFR response recorded at the frequency of the absent harmonic would be the result of noise or a possible DP; however there is no evidence of this in the study. When the harmonic is removed, the FFR response recorded at that frequency is either completely absent, or a small fluctuation that is likely the result of signal processing noise and not distortion.

Distortion products have been well documented in the neural responses to speech. Rickman and colleagues (1991) and Chertoff and colleagues (1992) validated the distortion product recorded using FFR at the frequency of 2F1-F2 (F1 and F2 being the primary tones of a two tone stimulus) as a true neural response and not an electrical artifact. Miller and colleagues (1997) detected neural responses to speech harmonic distortion products in the phase locking of the auditory nerve, while Krishnan (2002) recorded distortion products in the human FFR.

Pandya and Krishnan (2008) examined the characteristics of the 2F1-F2 distortion product. They elicited DP-FFRs using a variety of 2 tone complex tone burst stimuli, and presented the stimuli at 65, 75, 85, and 95 dB. It was determined that as the intensity of the stimulus decreased, all of the neural responses (F1, F2, and 2F1-F2) decreased. This is a reflection of the reduction in number of neurons involved in the processing of the stimulus; as the level of the stimulus increases, more neural components are recruited to process the stimulus. The amplitude of all components (F1, F2, and 2F1-F2) decreased with increasing stimulus frequency. As the stimulus frequency increases, there is a reduction in the ability of the neural elements to phase lock to the stimulus, thus decreasing the amplitude of the response recorded.

One of the possible hypotheses for the discrepancies between Krishnan’s findings (2002, 2008, 2009) and the present study in recording neural distortion product could be that distortion products are hidden by more complex stimuli. Krishnan used two tone burst stimuli while the present study used a vowel stimulus; with the vowel stimulus, there will be more activity along the cochlea and the central auditory system than with the tone bursts, thus potentially contaminating our ability to record a distortion product. If a distortion product is supposed to occur on a particular location on the basilar membrane, but the basilar membrane at that point is already occupied with a louder, more complex portion of the speech stimulus, the response to the distortion product could be masked by the complex stimulus.

A second hypothesis to explain our lack of DP-FFRs is generated by recent work from Aravamudhan and colleagues (2010) who looked at the effect of context on the speech FFR. It was found that the representation of the stimulus in the brainstem as measured by the FFR was greater when their syllabic stimulus (e.g. /ga/) was presented in isolation versus when it was preceded by another syllable (with a 50 ms silent gap between the two stimuli). While the purpose of the study was not to examine distortion products, it does provide evidence that FFR research completed with stimuli in isolation (Krishnan, 2002, 2008, 2009) cannot be directly compared to research completed with context, as we do see a reduction in the FFR following a preceding contextual stimulus. While previous studies on the FFR have documented distortion products, the majority of them presented isolated stimuli. The present study used multiple concatenated stimuli separated by a silent gap of 75 ms, so we could expect this reduction in the FFR as a result of the context, which may be sufficient to eliminate any distortion products.

Since we did not record any distortion products in the FFR to the vowel stimulus, it was of interest to determine if it was possible to measure a distortion product otoacoustic emission in the ear canal with our stimulus. A DPOAE recording system was designed to use the same stimulus, the same parameters/protocols as the FFR recording system, and the same fast Fourier transform analysis to determine the presence of DPOAEs. There was no evidence of distortion product otoacoustic emissions at any of the 13 harmonic frequencies; when they were absent from the acoustic stimulus, the measured response in the ear canal at that frequency disappeared. This suggests that it is not a lack of distortion in the brainstem response that we need to explain, but rather that distortion products are not present at any point along the auditory system for this particular stimulus.

4.3 COMPARISON OF BEHAVIORAL AND ELECTROPHYSIOLOGICAL FINDINGS

In general, the harmonic thresholds from the behavioral study follow a predictable pattern once the upward spread of masking is taken into consideration, and the amplitude of the FFR response to each harmonic can be predicted by their physical intensity in the stimulus. However, the results from the behavioral study do not predict the electrophysiological results. In fact, the correlation between FFR and pulsation threshold implies no causal relationship at all, when examined both across participants and within participants.

The most audible harmonic, which is the center harmonic of the first formant, is also the harmonic with the greatest FFR amplitude (H7) in the grand average; this suggests that the perceptual salience corresponds to the level of neural representation in the brainstem. However, an upward spread of masking effect was noted of F1 on adjacent higher frequency harmonics (H8, H9 being inaudible, and H10 and H11 having reduced audibility for a spectral peak). This effect was observed behaviorally, but the FFR responses do not support this finding as H8-H11 have significant FFR amplitudes. In a general sense, this suggests that the neural representation of the vowel stimulus in the brainstem (as measured by the FFR) does not predict the perception of the vowel harmonics, and the differences between the behavioral representation and the brainstem representation should be explored.

One of the questions that arises out of this finding that the FFR does not predict the pulsation threshold task is the source of the frequency-following response and the basis for the pulsation threshold task. It has generally been agreed upon that the FFR is generated by a neural population in the rostral brainstem (Glaser *et al.,* 1976). Using ablation studies and developmental studies, it has been localized to the inferior colliculus, the lateral lemniscus, and the cochlear nucleus, and in humans, scalp-FFRs are absent in participants with selective lesions of the inferior colliculus (see Chandraskaran and Kraus (2010) for a review). Smith and colleagues (1975) induced a selective amplitude reduction of the FFR in cats following cryogenic treatment of the inferior colliculus, and the amplitude recovered once the inferior colliculus was warmed. These findings illustrate that the FFR is a response due to the phase locking of upper brainstem neurons.

Kielson and colleagues (1997) examined the cells of the ventral cochlear nucleus and evaluated their role in segregating competing speech sounds. By isolating the cat VCN, they classified the single units into primary like cells or chopper cells based on firing patterns. Two syllables were played simultaneously to the cat, and the response of the single unit to the syllables was recorded. The chopper cells are of interest with regards to this study; the chopper cells phase lock to the fundamental frequency of the vowel, and provide a spectral representation of the stimulus based on the tonotopic mapping of the VCN cells by best frequency. Thus, a single chopper cell represented the pitch of the sound by phase locking, but the spectrum of the sound was represented across a population of the VCN. This hypothesized model of segregating speech sounds in the ventral cochlear nucleus by chopper cells may account for the mismatch between the electrophysiology and the behavioral task of this study, in that the neural encoding that is relevant for speech perception (the rate-place code of the vowel spectrum) may not be reflected in the spectrum of the FFR.

The FFR to a pure tone reflects the phase locking of a population of neurons in the brainstem; these neurons are responding to activity at a particular frequency, not necessarily activity at a particular place on the basilar membrane. Thus, the temporal representation of the sound in the cochlea is reflected by the activity measured as the FFR. However, this temporal representation does not seem to have an effect on our perception of the vowel stimulus; harmonics that are represented temporally (H8-H9) in the brainstem are not audible in the behavioral task. This suggests that although the harmonic is represented temporally, the temporal representation is not useful for the detection of harmonics behaviorally.

The FFR reflects neural synchrony, so it relates to information that is temporally coded, however the place specificity of the FFR also needs to be discussed. Yamada and colleagues (1978) found that FFRs elicited by 500 Hz tone bursts presented at low intensities reflected activity at a restricted region of the apical end of the cochlea.

However, place specificity has not been successfully found for moderate to high intensities. Krishnan (1992) used a tone on tone forward masking paradigm to evaluate the place specificity of the 500 Hz tone burst FFR elicited at moderate intensity. It was found that the FFR to 500 Hz was the product of upward spread (towards the base) of excitation, so the FFR has a basalward bias from the probe frequency. For example, the 500 Hz FFR was generated from a restricted apical region at 1000 Hz. These findings suggest that the neurons that are responsible for the phase locking to 500 Hz to generate the 500 Hz FFR are not found in the corresponding “500 Hz place” in the tonotopic map. Thus, the FFR would not necessarily be affected by peripheral mechanisms such as simultaneous masking.

This study suggests that we get temporal phase locking to sounds we can’t perceive; this information exists in the system, but we don’t have direct conscious access to it for sound identity. In the cochlea and central auditory system, the neurons with characteristic frequencies near H8 may not be actively encoding the eighth harmonic because they are masked by energy at H7. Thus, H8 becomes inaudible. However, an FFR may still occur to H8 because phase locking may be driven by neurons that are not masked (i.e., in other regions of the cochlea). Phase-locking to H8 does not have to originate in the neurons most responsive to H8; H8 could generate enough movement on the basilar membrane to force some of the neurons with characteristic frequencies near H9 to phase lock to H8. This would cause a temporal coding to H8, but no place code to H8 since the neurons whose CFs correspond to H8 are masked by H7. Due to the lack of activity in the central auditory system to H8 neurons, perception of H8 is reduced. This hypothesis suggests that our perception of harmonics does not depend on the temporal locking by the brainstem to those harmonics, but rather a rate-place code of the central auditory system that is tonotopically mapped. The basic assumption of the pulsation threshold is that it approximates the tuning curves of inner hair cells; this would suggest that the pulsation threshold results are a better representation of place coding. It is not unreasonable to hypothesize that the temporal code established by the FFR is capturing an inaudible portion of the spectrum, and that the results of the pulsation threshold task are not predicted by the FFR because the audibility of the harmonics is derived from a rate-place code rather than a purely temporal code.

The act of listening to a specific harmonic in a harmonic complex is an abnormal circumstance; we normally group harmonics together rather than try to separate them out. The overall representation of the harmonic complex provides us with the sensation of timbre. In difficult listening situations or while listening to complex stimuli, both temporal and place coding may be important.

There are a number of reasons to doubt that the FFR to speech harmonics has any relationship with harmonic or formant audibility. First, we can hear tones above 1500 Hz but we cannot easily record FFRs above this frequency due to the low-pass filter function of the brainstem. Interestingly, perception of pitch does decrease after 1500 Hz, which offers further evidence that temporal brainstem encoding is important for pitch coding. Also, in a number of vowels, the second formant occurs at a frequency that is too high to generate an FFR, but is nonetheless critical for speech understanding.

The temporal representation by the brainstem has been found to be important for pitch perception. A subcortical determination of pitch has been suggested based on findings relating FFR and behavioral ratings of consonance/dissonance. Bidelman and

Krishnan (2009) asked non-musicians to select the more pleasant sounding tone dyad in a paired comparison task. They also recorded the frequency-following response to the tone dyads using dichotic presentation. It was found that FFRs in response to consonant dyad intervals were more robust than the FFRs to the dissonant stimuli. Furthermore, behavioral consonance ratings were predicted by the neural pitch salience (r=0.81), suggesting that a listener's tonal percept could be predicted by the subcortical activity. Consonant pitch intervals were judged as more pleasant by the listeners, and a more robust FFR was recorded to the consonant pitch intervals. These results suggest that the perception of musical pitch may arise from the temporal processing at subcortical levels. While the final formation of a musical tone percept may occur at the cortical level, these findings suggest that the synchronous responses in the brainstem play an important role in pitch perception. Thus, there is evidence for a relationship between the brainstem activity and perception of pitch in humans. The present study has suggested that there is no predictable relationship between the temporal encoding of speech in the brainstem and perception of harmonics in humans.

4.4 EXISTING MODELS OF VOWEL PERCEPTION

In light of the finding that temporal representation is not sufficient for assigning sound identity to a component of a vowel stimulus, it is of interest to evaluate the existing models of vowel perception. The simplest models of vowel perception are based on representation of the formant frequency alone. It was thought that the first two formants of a vowel were sufficient for vowel perception (Peterson and Barney, 1952), and that other properties of the vowel spectrum (formant amplitude, bandwith, spectral tilt) were unimportant in vowel identification (Klatt, 1982). However, it has been suggested that these models are exceedingly basic, as we are unable to predict the identification of a vowel based on formant frequency alone (see Rosner and Pickering, 1994 for a review).

Certain authors propose a model of vowel perception based on a broad representation of spectral properties of the vowel stimulus (Blandon and Lindblom, 1981; Ito *et al.,* 2001). Zahorian and Jagharghi (1993) were able to use algorithms of the spectral shape parameters to classify vowels more accurately than the classification that occurred based on formant frequencies alone. Spectral shape models ignore the importance of formant frequencies entirely.

An intermediate model suggests that a preliminary analysis of the vowel occurs on the basis of formant and fundamental frequencies, and a secondary analysis results in the integration of the spectral peaks over a broad frequency range. The final percept of the vowel is based on perceived peaks, but these peaks may not correspond directly to the formants. In a study of vowel identity, participants were asked to adjust the frequency of F1 and F2 of a vowel stimulus until it matched the prototype vowel stimulus (Carlson, 1970). Participants matched the F1 of the test stimulus and the prototype stimulus very accurately. However, the general trend with the F2 matching was that participants matched the F2 frequency of the test vowel to the F3 frequency of the prototype. Carlson and colleagues (1970, 1975) suggested that formant peaks closer than 3-3.5 bark are merged into the perception of a single spectral peak instead of two formant peaks. When these "superformants" are formed, other spectral properties such as formant amplitude can alter the perception of the vowel identity by influencing the local spectral balance, causing a shift in the perceived spectral peak. This smearing of narrowly spaced formants results in a more broadband processing of the vowel spectrum, whereby formant amplitude can affect the perception of the vowel under certain conditions. It has been established that changes in formant amplitude can be perceived independently of changes in formant frequency (Bernstein, 1981). However, other authors (Assmann, 1991) have found no effect for formant amplitude on perceived vowel identity, and claim that the manipulation of formant amplitude does not have an effect on the phonetic quality of the vowel (Lindblom *et al*., 2009).

Carlson and colleagues (1970, 1975) were the first to propose the model of the superformants, suggesting a higher level of processing beyond the peripheral models; however, there was no evidence to support this finding. This study is a demonstration of a vowel effect that cannot be understood completely from the brainstem recordings. A potential theory is that there is a rate-place code for vowel perception rather than a purely temporal code of the vowel spectrum. Perception of formants requires that the two formants be separated by a minimum of 100 Hz, however, the human frequency difference limen is significantly smaller than 100 Hz. This mismatch may suggest that the time coding of the formants does not account for the complexity of vowel perception. At conversational speech levels (60-70 dB), the individual firing rates of mid- and high-spontaneous rate fibers saturated, resulting in poor frequency selectivity due to flat frequency-rate curves (Blackburn and Sachs, 1990; Recio and Rhode, 2000). This effect has been shown to be the result of both the limited dynamic range of the fibers, as well as two tone suppression occurring within the vowel (Schalk and Sachs, 1979). Low spontaneous rate fibers were shown to encode spectral peaks up to 70-80 dB (Sachs and Young, 1979). These findings suggest that a representation of formant frequencies based on individual auditory nerve fiber firing rates is inadequate for the transmission of information necessary for vowel perception at higher speech levels. However, Recio and colleagues (2002) found that when the spectral composition of the vowel was adjusted to match the human cochlear distance rather than frequency separation along the animal basilar membrane, rate based encoding of speech stimuli provided more information that previously documented, even at high speech levels. While further research is required to be certain, these findings suggest that more information can be encoded in the firing rate of neurons than previously documented.

Kielson *et al*. (1997) discovered that a single chopper cell in the ventral cochlear nucleus represented the pitch of a speech sound by phase locking, but the spectrum of the speech sound was represented across the population of the chopper cells in the VCN. The chopper cells were shown to encode the harmonic spectrum using spatial representation, and the pitch of the sound using temporal representation. This is a cellular example of both a rate and a place code of a speech stimulus, suggesting that speech may be exceedingly complicated for a solely peripheral processing strategy. The electrophysiological results of this study show that there is more temporal information available in the central auditory system than one might expect on the basis of the behavioral responses. The finding that this temporal information cannot be perceived suggests that purely temporal models of vowel perception that rely solely on the temporal representation of formant frequencies may be too basic to account for speech decoding.

The temporal representation of the sound in the central auditory system may be important for providing additional information that facilitates speech understanding in complex listening environments. Smith and colleagues (2002) used auditory chimera stimuli that were composed of the envelope of one speech sample and the harmonics of a different speech sample to demonstrate that the perception of speech was dominated by the envelope. Thus, the listeners perceived the sentence whose envelope structure was contained in the auditory chimera stimulus and not the sentence that matched the harmonic structure. However, melody-melody chimeras showed that the perception of pitch was dominated by the sound whose harmonics were contained in the chimera. Finally, the sound of a speech-speech chimera was heard at the location determined by the harmonic structure. These findings suggest that while the temporal representation of the harmonics in the central auditory system may not be facilitating speech perception, they contribute to the act of listening by encoding pitch information and sound location in auditory space.

4.5 CLINICAL IMPLICATIONS

There is a great need to have a quick clinical test of overall speech decoding that does not require a verbal response and thus can be used with infants or difficult to test clients, however the FFR should not be used to predict which aspects of the speech signal are audible. While it does give us information about the representation of temporal information in the brainstem, it cannot be used to provide us information about hearing thresholds or speech audibility per se. The FFR is not place specific; it is only generated robustly at high levels that are associated with poor cochlear place specificity, and there is no guarantee that a frequency in the response is generated at the cochlear place corresponding to that frequency. In contrast, the behavioral results provide information about which parts of the speech signal are audible, but no information on the temporal representation of the stimulus. These two measures are thus complementary, and provide different types of information; neither should be used to replace the other. The lack of relationship between the behavioral response and the electrophysiology recordings also serves as a reminder that clinicians and researchers need to be cautious in the interpretation of neural responses.

4.6 POTENTIAL ISSUES WITH THE STUDY

An important question to address is whether we measured an excessive amount of electrical current artifact or electrical noise, due to the unexpected patterns we observed in our recordings (lack of DP-FFRs). There are a number of reasons why we are confident that our measurements reflect neural activity and not electrical noise. First, there were substantial differences between participants, despite a constant setup. Variability in brainstem recordings is expected due to a number of factors such as age, sex, hearing acuity, etc so the variability in the recordings reflect this fact. If it was electromagnetic radiation, the responses would be very constant across individuals and we would not see the same variability in response.

Furthermore, the amplitude of the response attenuates with increasing frequency, reflecting the low pass filter of the brainstem. Since the FFR recording reflects the phase-locking of a population of neurons, as the frequency of interest increases, the overall FFR amplitude decreases, reflecting the upper limits of neuronal phase locking abilities. The overall attenuation of the FFR in the present study suggests that we are recording the phase-locked activity of a population of neurons, and not electrical noise. Also, if it was electromagnetic radiation, we would expect the overall frequency response to be flat, whereas there are distinct low frequency components recorded in the present study.

To further examine this possibility, a participant was tested in the identical setup to the experiment, but the recordings were made with the tube of the earphone encased in putty, thus they were not receiving any acoustical stimulation. The participant was tested in a double-walled sound booth, but to ensure no external sounds were heard, the participant wore earplugs bilaterally. If the FFR we recorded reflected electrical noise, the energy in the FFR with and without the acoustic stimulus would be similar. There was no FFR in response to the recordings made without acoustic stimulation, and the only frequencies that had peaks in amplitude were multiples of 60 Hz, reflecting line noise. After the trial without acoustic stimulation was completed, the earplugs were removed and the insert plug was placed in the ear, and the participant was tested with acoustic stimulation to ensure the FFR was recorded in response to the stimulus. There was a robust FFR recorded to the harmonics of the stimulus, indicating that the data of this study accurately reflects the brainstem response and not noise. The lack of DP-FFRs is an unexpected finding, but it cannot be explained by a recording error.

4.7 FUTURE DIRECTIONS

One study that could verify the hypothesis that the FFR to inaudible frequencies of the spectrum (inaudible due to masking by lower frequency components) is generated by neighbouring higher-frequency neurons would be to perform a similar task but with high frequency masking. The high frequency masking would occupy the higher frequency neurons that are hypothesized to be phase locking to the behaviourally inaudible lower frequencies; if the FFR to a masked harmonic disappears with masking, this is evidence that it is higher frequency neurons generating the FFR at the inaudible frequency.

4.8 GENERAL CONCLUSIONS

This study attempted to determine the audibility of individual harmonics in a vowel stimulus, and whether the ability to hear a specific harmonic relates to the degree of neural synchronization to that harmonic. The individual harmonics were encoded in the brainstem, as determined using FFR, and the individual harmonics were audible to the individual, as determined using the pulsation threshold technique. Certain harmonics were subject to the expected masking effects in the periphery, however, both the FFR response to the harmonics and the behavioral response to the harmonics followed an expected pattern. The study also attempted to establish a relationship between the degree of brainstem representation of the harmonic and the perceptual salience of the harmonic, and there was no relationship observed. The FFR reflects temporal coding, while it appears that harmonic audibility is the result of a rate-place code; temporal phase locking was observed to frequencies that were inaudible behaviorally. In general, this suggests that the FFR should not be used as a test of speech harmonic audibility since the temporal information present in the FFR cannot be used to decode a vowel stimulus without the place code information.

BIBLIOGRAPHY

Aaltonen, O. (1985). The effect of relative amplitude levels of F2 and F3 on the categorization of synthetic vowels. *Journal of Phonetics,* 13, 1–9.

Aiken, S.J., & Picton, T.W. (2006). Envelope following responses to natural vowels. *Audiology and Neurotology 11(4),* 213–232.

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

Aravamudhan, R., Carbonell, K., and Lotto, A. (2010). Presence of preceding sound affects the neural representation of speech sounds: Frequency following response data. *Journal of the Acoustical Society of America, 128(4)*, 2322-2322.

Assmann, P. F. (1991). The perception of back vowels: Center of gravity hypothesis. *Journal of Experimental Psychology,* 43A, 423–448.

Bernstein, J. (1981). Formant-based representation of auditory similarity among vowel-like sounds. *Journal of the Acoustical Society of America,* 69, 1132–1144.

Bidelman, G, & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience,* 29(42), 13165-13171.

Blackburn, C. C., and Sachs, M. B. (1990). The representation of the steady-state vowel sound /ε/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *Journal of Neurophysiology*, 63, 1191–1212.

Bladon, R. A. W., & Lindblom, B. (1981). Modeling the judgement of vowel quality differences. *Journal of the Acoustical Society of America*. 69. 1414–1422.

Bregman, A. (1993). Auditory scene analysis: hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: the cognitive psychology of human audition* (10-36). Oxford University Press.

Brownell, W.E. (1990). Outer hair cell electromotility and otoacoustic emissions. *Ear and Hearing,* 11, 82-92.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch, and vowel perception. *Auditory Analysis and Perception of Speech*, Academic, London, pp. 55–82.

Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. Speech Transm. Lab. Q. Prog. Status Rep. 11, 19–35.

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

Chertoff, ME., Hecox, KE., & Goldstein R. (1992). Auditory distortion products measured with averaged auditory evoked potentials. *Journal of Speech and Hearing Research,*35, 157–166.

Dajani, H., Purcell, D., Wong, W., Kunov, & H., Picton T. (2005a) Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52, 1614-1618.

Dimitrijevic, A., John, M.S., & Picton, T.W. (2004) Auditory steady-state responses and word recognition scores in normal-hearing and hearing-Impaired adults. *Ear and Hearing,* 25, 68-84.

Gelfand, S. (2004). *Hearing- An Introduction to Psychological and Physiological Acoustics* 4th Ed. New York: Marcel Dekker.

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: 25–32.

Houtgast, T. (1974). Auditory analysis of vowel-like sounds. *Acoustica, 31*, 320-324.

Huis in"t Veld, F., Osterhammel, P., & Terkildsen, K. (1977). The frequency selectivity of the 500 Hz frequency following response. *Scandinavian Audiology,* 6, 35-42.

Ito, M., Tsuchida, J., & Yano, M. (2001). On the effectiveness of whole spectral shape for vowel perception. *Journal of the Acoustical Society of America,*110, 1141–1149.

Kakusho, O., Hirato, H., & Kato, K. (1971). Some experiments of vowel perception by harmonic synthesizer. *Acoustica* 24, 179–190.

Kiefte, M., Enright, T., & Marshall, L. (2010). The role of formant amplitude in the perception of /i/ and /u/. *Journal of the Acoustical Society of America,* 127 (4), 2611-2621.

Kielson, S.E., Richards, V.M., Wyman, B.T., Young, E.D. (1997). The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *Journal of the Acoustical Society of America,* 102, 1056-1071.

Klatt, D. H. (1979). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971–995.

Klatt, D. H. (1982). Prediction of perceived phonetic distance from criticalband spectra: A first step. *Acoustic Speech and Signal Processing;* 7, 1278–1281.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research 166 (1–2),* 192–201.

Krishnan, A., Xu, Y., Gandour, J.T., & Cariani, P.A. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research, 168*, 1-12.

Krishnan A, Gandour JT (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language,* 110, 135–148.

Levit, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

Liberman, A.M., Delattre, P.C., Cooper, F.S., & Gerstman, L.J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs 68 (8)*, 1-13.

Liberman, M.C., Zuo, J, Guinan, J.J. (2004). Otoacoustic emissions without somatic motility: can stereocilia mechanics drive the mammalian cochlea? *Journal of the Acoustical Society of America,* 116, 1649-1655.

Lindblom, B., Diehl, R., &Creeger, C. (2009). Do 'dominant frequencies' explain the listener's response to formant and spectrum shape variations?. *Speech Community,* 51, 622–629.

Macintosh, S., & Kiefte, M. (2005). Vowel masked audiograms. Unpublished master's project, Dalhousie University.

Miller, R.L., Schilling, J.R., Franck, K.R., Young, E.D. (1997). Effect of acoustic trauma on the representation of the vowel "eh" in cat auditory nerve fibers. *Journal of the Acoustical Society of America,* 101, 3602-3616.

Moore, B.C.J., & Glasberg, B.R. (1983). Masking patterns for synthetic vowels in simultaneous and forward masking. *Journal of the Acoustical Society of America, 73*, 906-917.

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*, 665–667

Nearey, T. M., & Levitt, A. G. (1974). Evidence for spectral fusion in dichotic release from upward spread of masking, Haskins Laboratories: Status Rep. Speech Res. SR-39, 81–89.

Pandya, P., & Krishnan, A. (2004). Human frequency-following response correlates of the distortion product at 2F1-F2. *Journal of the American Academy of Audiology*, 15, 184-197.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America,* 24*,* 118-127.

Plyler, P., & Ananthanarayan, A.K. (2001). Human frequency-following responses: representation of second formant transitions in normal and hearing-impaired listeners. *Journal of the American Academy of Audiology,* 12, 523-533.

Purcell, D.W., Ross, B., Picton, T.W., Pantev, C. (2007). Cortical responses to the 2f1-f2 combination tone measured indirectly using magnetoencephalography. *Journal of the Acoustical Society of America,* 122, 992-1003.

Recio, A., Rhode, W., Kiefte, M., & Kleunder, K. (2002). Responses to cochlear normalized speech stimuli in the auditory nerve of cat. *Journal of the Acoustical Society of America,* 111 (5), 2213-2218.

Rickman, MD., Chertoff, ME., & Hecox, KE. (1991). Electrophysiological evidence of nonlinear distortion products to two-tone stimuli. *Journal of the Acoustical Society of America*, 89,2818–2826.

Rosner, B. S., & Pickering, J. B. (**1994**). *Vowel Perception and Production*. Oxford University Press, Oxford.

Schalk, T. B., and Sachs, M. B. (1979). Nonlinearities in auditory-nerve fiber responses to bandlimited noise. *Journal of the Acoustical Society of America,* 67, 903–913.

Smith, Z., Delgutte, B., & Oxenham, A. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87-90.

Smith, J.C., Marsh, J.T., Brown, W.S. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology,* 39, 465-472.

Thurlow, W.R., & Efner, L.F. (1959). Continuity effects with alternately sounding tones. *Journal of the Acoustical Society of America,* 31, 1337-1339.

Vogten, L.L.M. (1974). Pure-tone masking: A new result from a new method. In E. Zwicker & E. Terdhart (Eds.), *Facts and Models in Hearing*. Berlin: Springer-Verlag.

Warren, R.M., Obusek, C.J., and Ackroff, J.M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science,* 176, 1149-1151.

Yamada, O., Yamane, H., & Kodera, K. (1977). Simultaneous recordings of the brain stem response and the frequency-following response to low-frequency tone. *Electroencephalography and Clinical Neurophysiology,* 43, 362-370.

Young, ED., Sachs, MB. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *Journal of the Acoustical Society of America*, 66, 1381–1403.

Zahorian, S. A., & Jagharghi, A. J (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America,* 94, 1966-1982.

APPENDIX

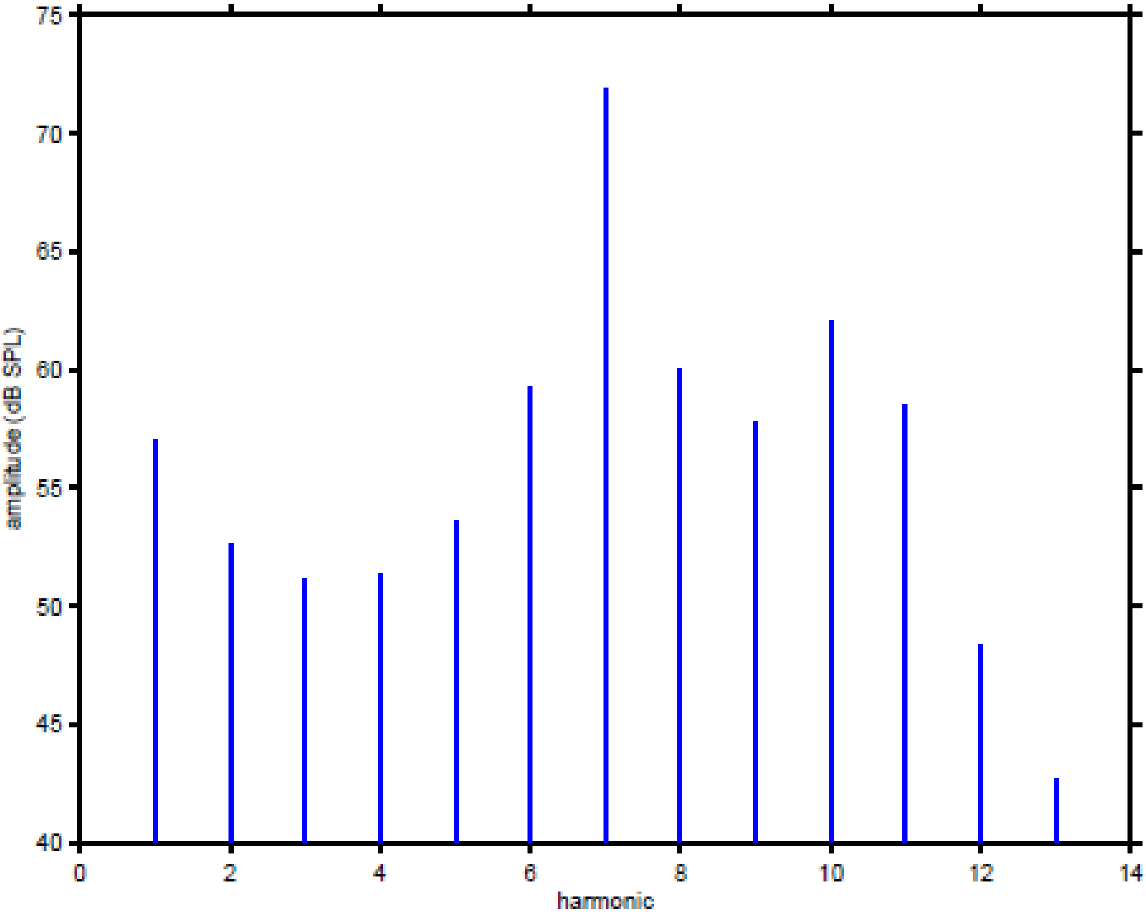


Figure 1. Harmonic spectrum of the synthetic vowel /ɒ/ created using harmonic frequencies of a live male voice. The amplitude of harmonics 1-13 are represented in dB SPL.

Harmonic Number	Frequency (Hz)
1(F0)	114.084
2	228.168
3	342.252
4	456.336
5	570.420
6	684.504
7	798.588
8	912.672
9	1026.756
10	1140.84
11	1254.924
12	1369.008
13	1483.092
Formant	Frequency
F1	790
F2	1195
F3	2736

Table 1. The harmonic and formant frequencies for the synthetic vowel /ɒ/ based on the live male voice recording.

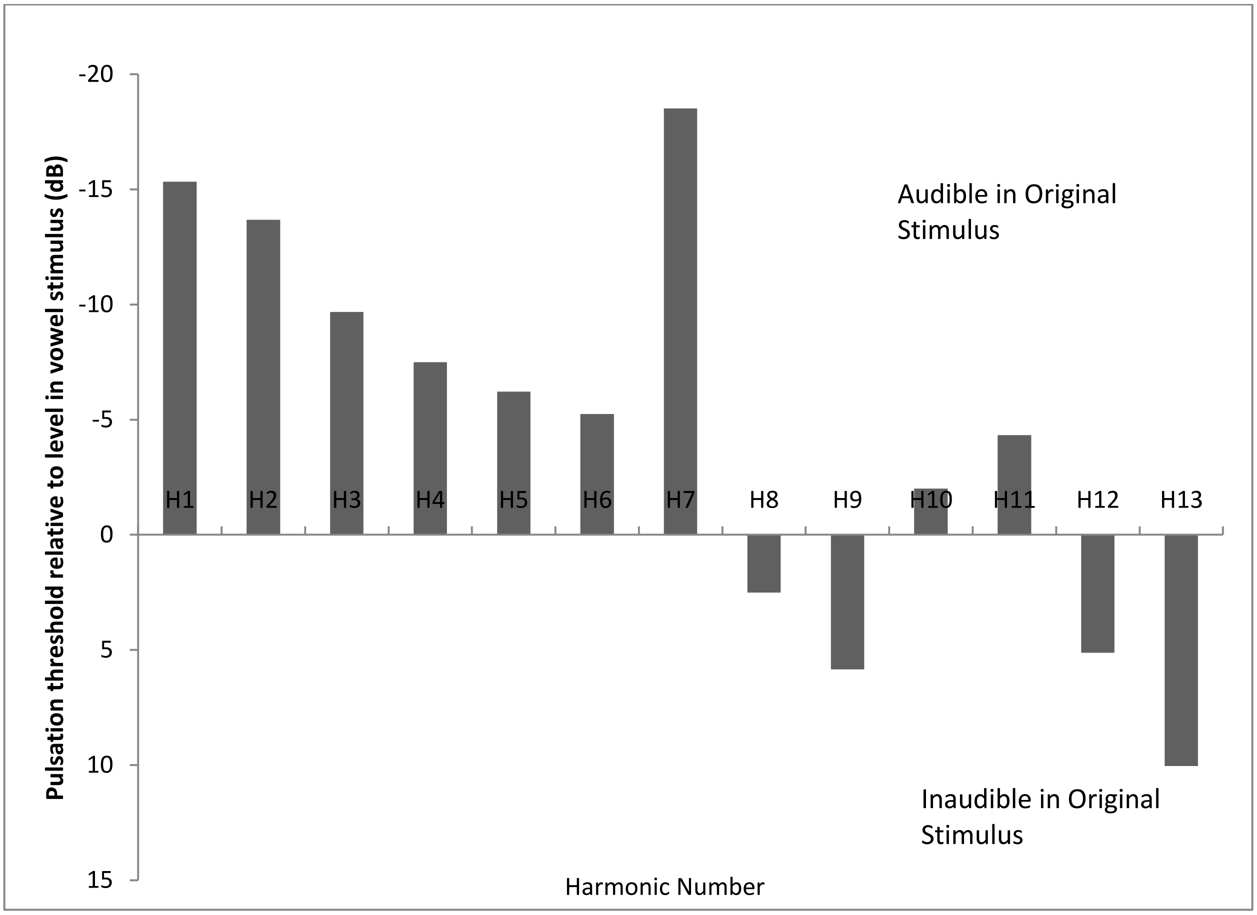


Figure 2. Grand average pulsation thresholds across harmonics in relative dB. A negative threshold indicates that the harmonic has high perceptual salience, and is audible in the vowel stimulus. A positive threshold indicates that the harmonic has no perceptual salience, and is inaudible in the vowel stimulus. H7 is the most audible harmonic; H7 is the harmonic frequency centered in the first formant of the vowel.

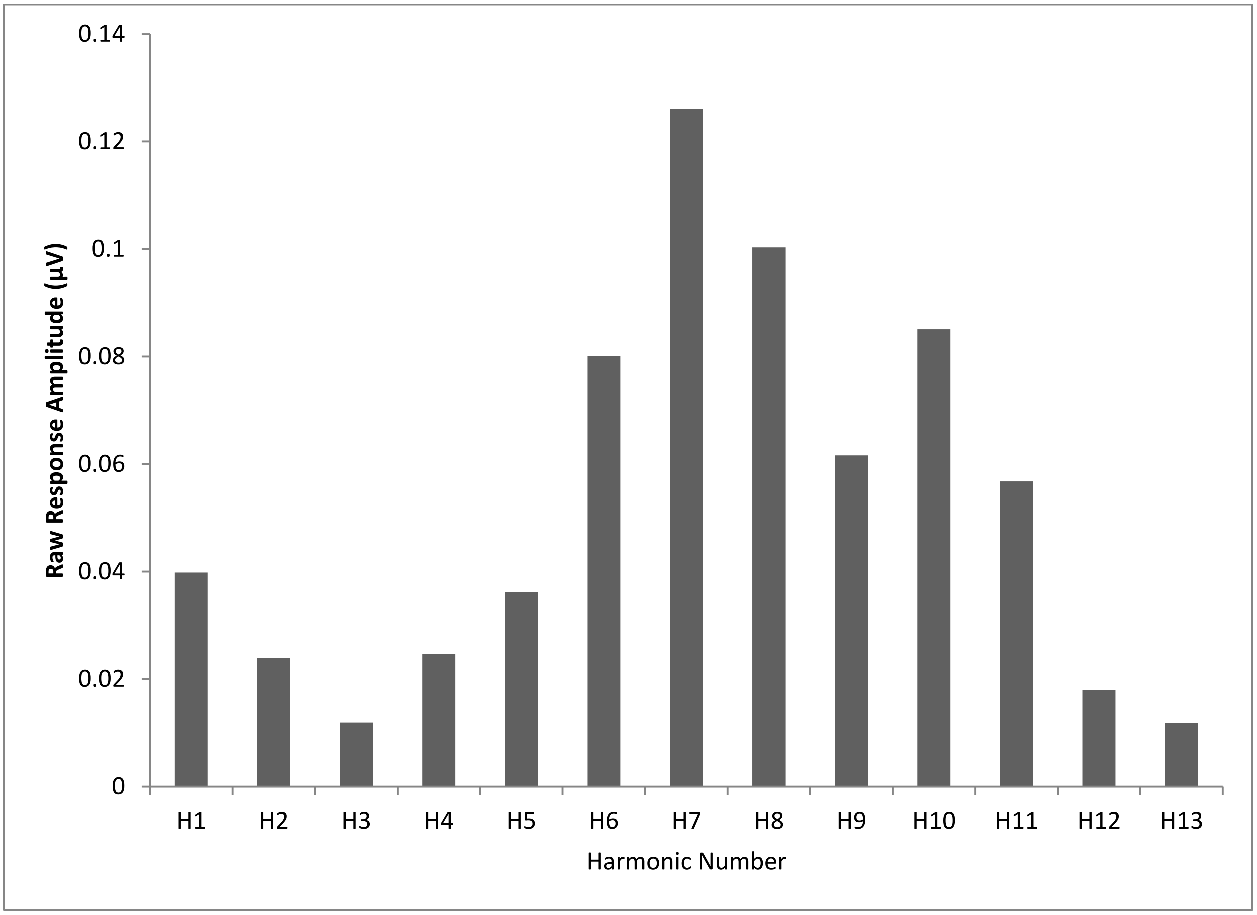


Figure 3. Grand average FFR amplitudes across harmonics, obtained by Fast Fourier Transform. The raw amplitudes (μV) at each harmonic for the full stimulus include the FFR to the harmonic, any distortion products, and noise. The largest FFR responses are found at H7 and H8 (the harmonic frequencies contained by F1) and H10, a frequency of the second formant. There is less of a response at the higher frequencies, which is expected due to the low pass function of the brainstem.

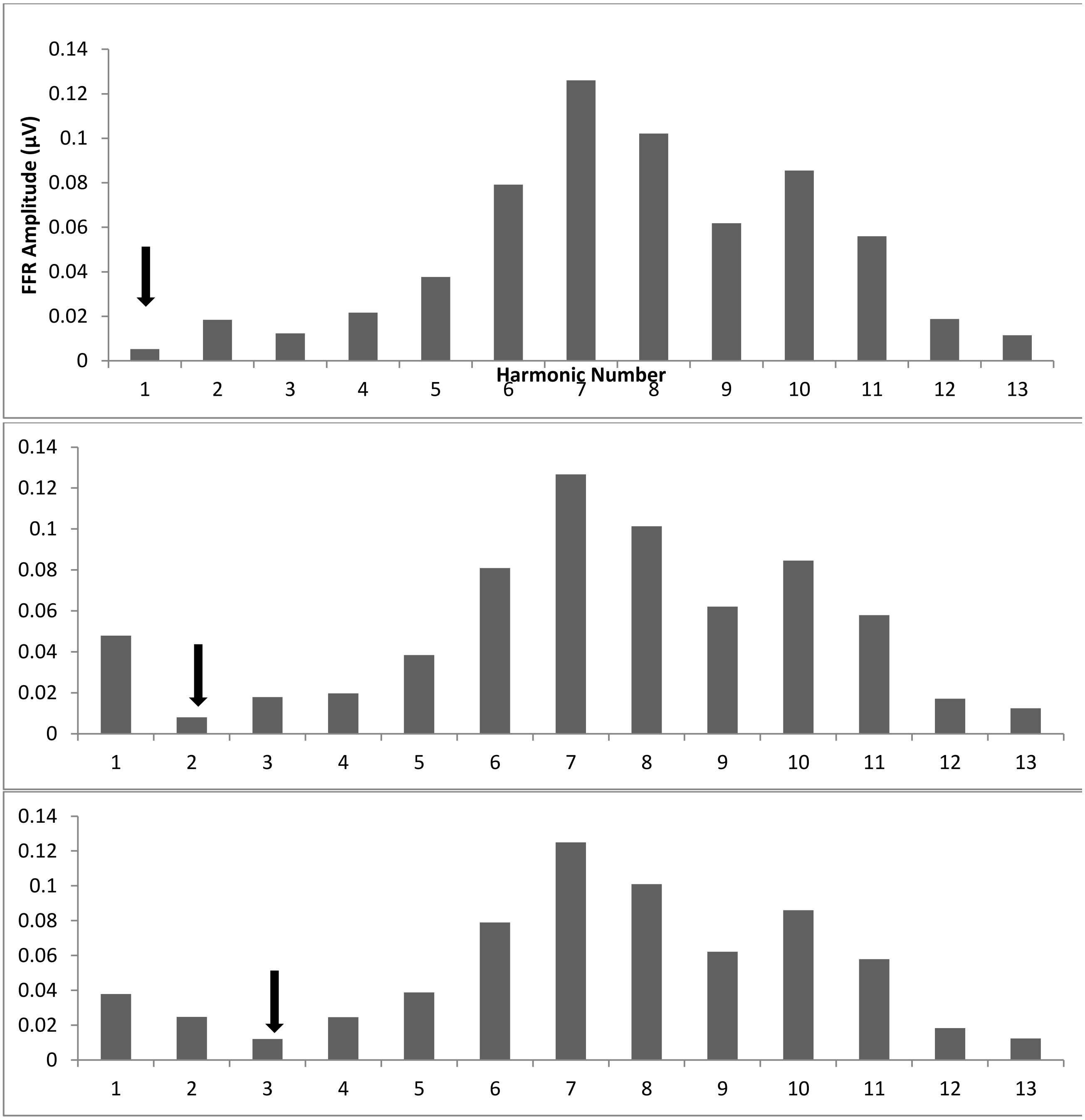


Figure 4. Grand average FFR amplitudes (µV) to vowel stimulus with a single harmonic removed from the stimulus. Arrow represents the missing harmonic. Figure continued on next page.

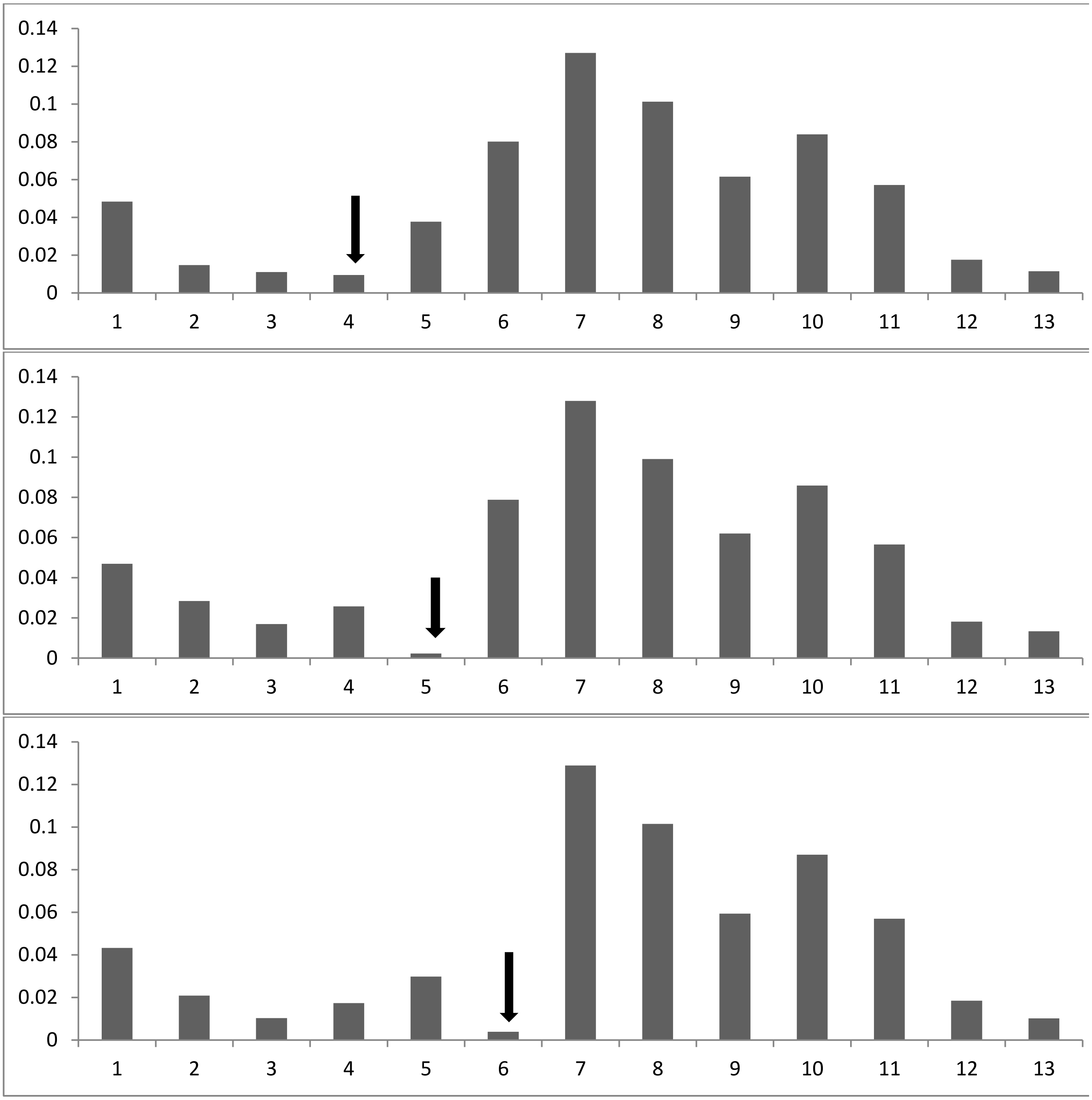


Figure 4 continued. Grand average FFR amplitudes (μV) to vowel stimulus with a single harmonic removed from the stimulus. Arrow represents the missing harmonic. Figure continued on next page.

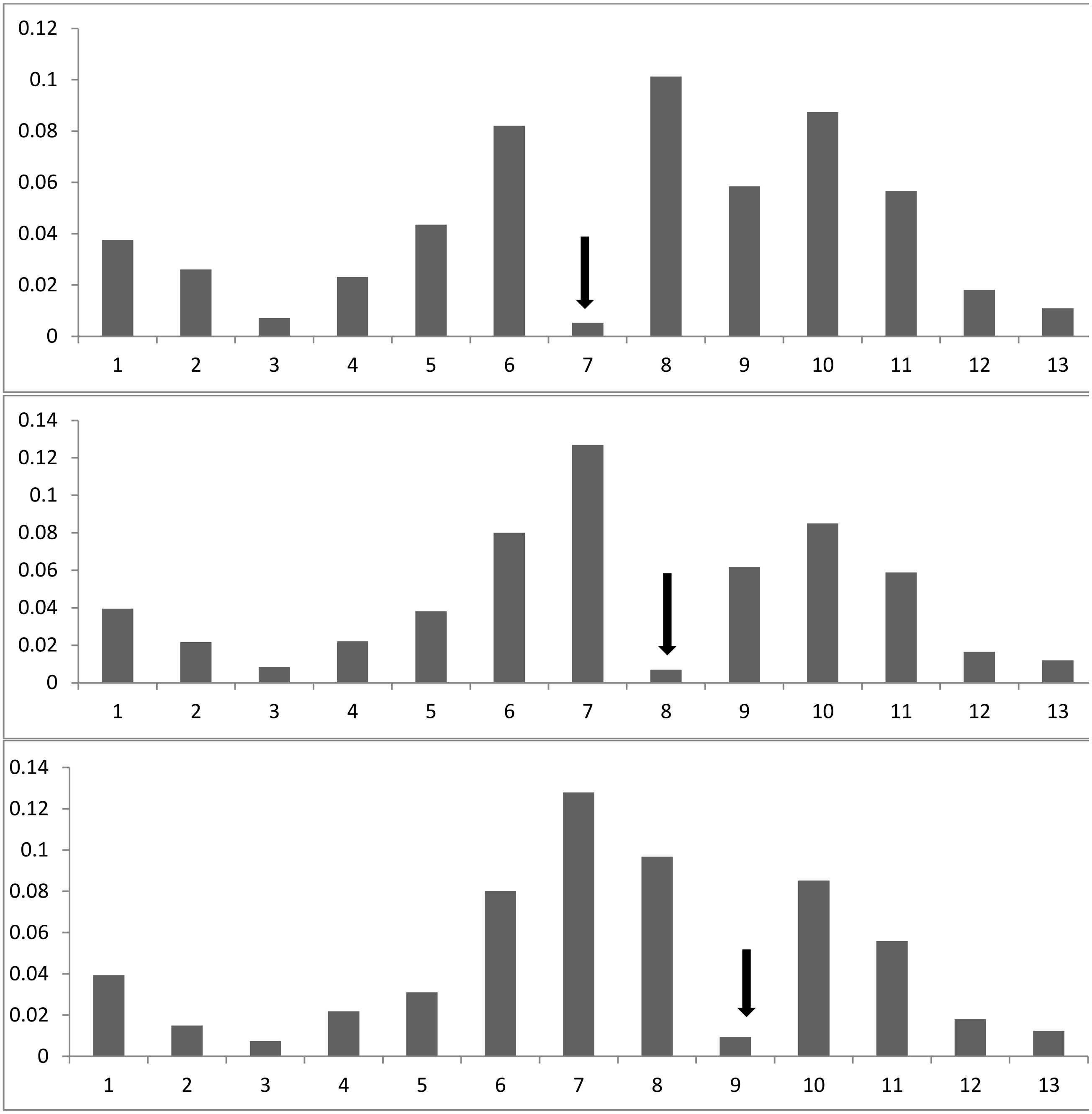


Figure 4 continued. Grand average FFR amplitudes (μV) to vowel stimulus with a single harmonic removed from the stimulus. Arrow represents the missing harmonic. Figure continued on next page.

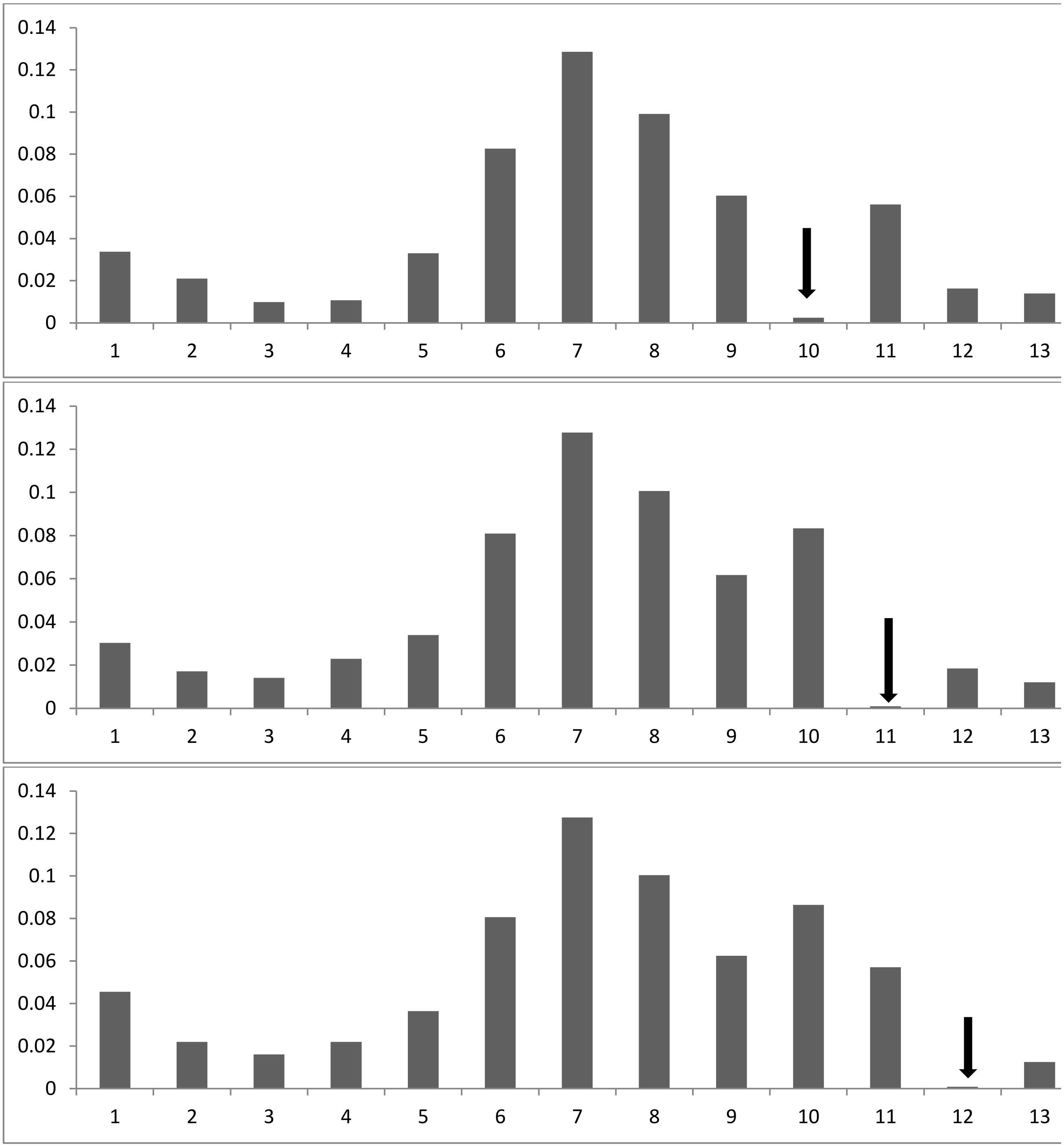


Figure 4 continued. Grand average FFR amplitudes (μV) to vowel stimulus with a single harmonic removed from the stimulus. Arrow represents the missing harmonic. Figure continued on next page.

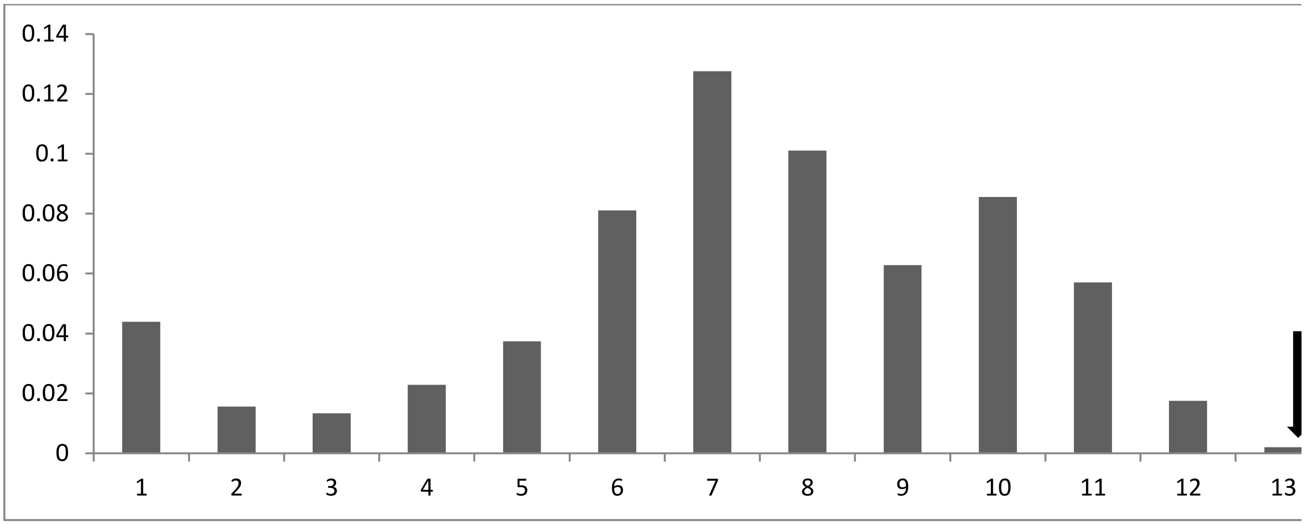


Figure 4 continued. Grand average FFR amplitudes (μV) to vowel stimulus with a single harmonic removed from the stimulus. Arrow represents the missing harmonic. This set of graphs demonstrates that the response to a single harmonic is almost entirely related to the energy at that harmonic in the stimulus; once that harmonic is removed from the stimulus the response is greatly reduced.

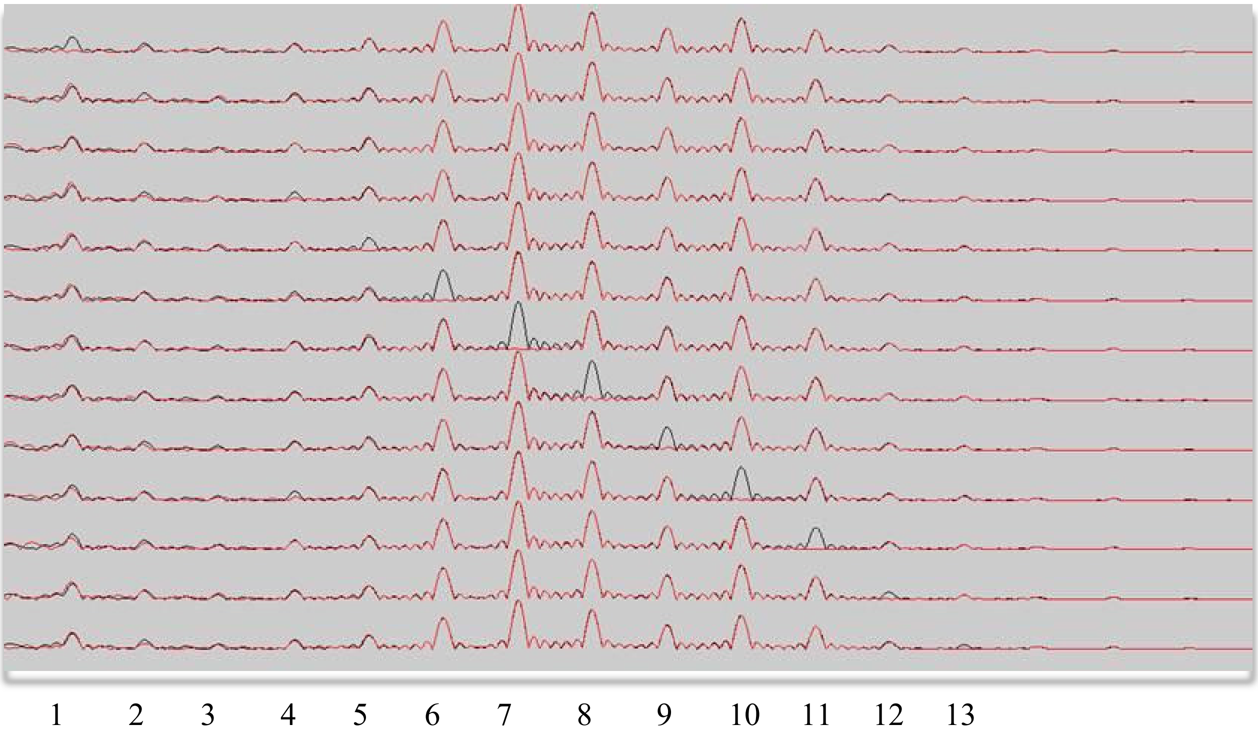


Figure 5. Grand average FFR response spectra. The black line is the response to the full stimulus, and the red line is the response to the stimulus with a single harmonic missing. The missing harmonic begins with 1 on the top, down to 13 on the bottom of the graph. In general, the energy at each harmonic in the response appears to be almost entirely dependent on the energy at the harmonic in the stimulus.

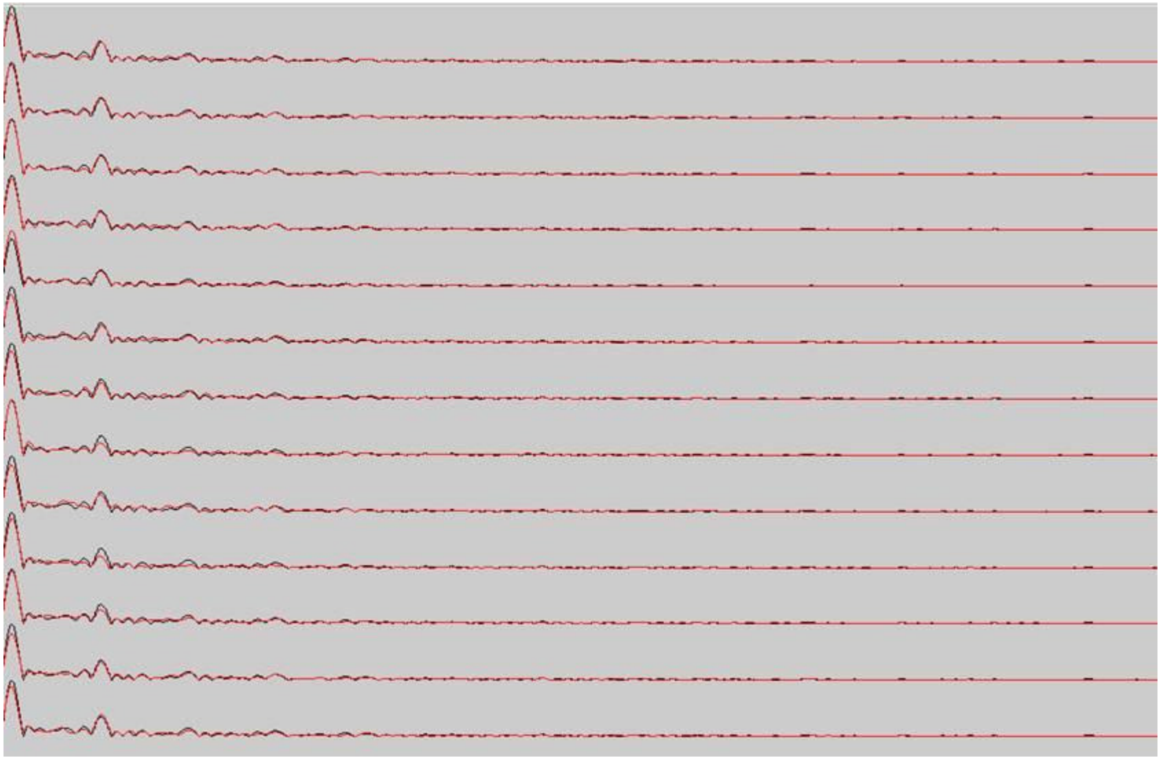


Figure 6. The alternating polarity (envelope response) grand average across harmonics shows a response at the fundamental frequency and the lower harmonics. The black line is the envelope response to the full stimulus. The red line is the envelope response to the stimulus missing a single harmonic (from 1 harmonic missing (top tracing) to 13 harmonic missing (bottom tracing)). This energy is reduced slightly when the first harmonic is removed, and also when the harmonics near the formant peaks are removed.

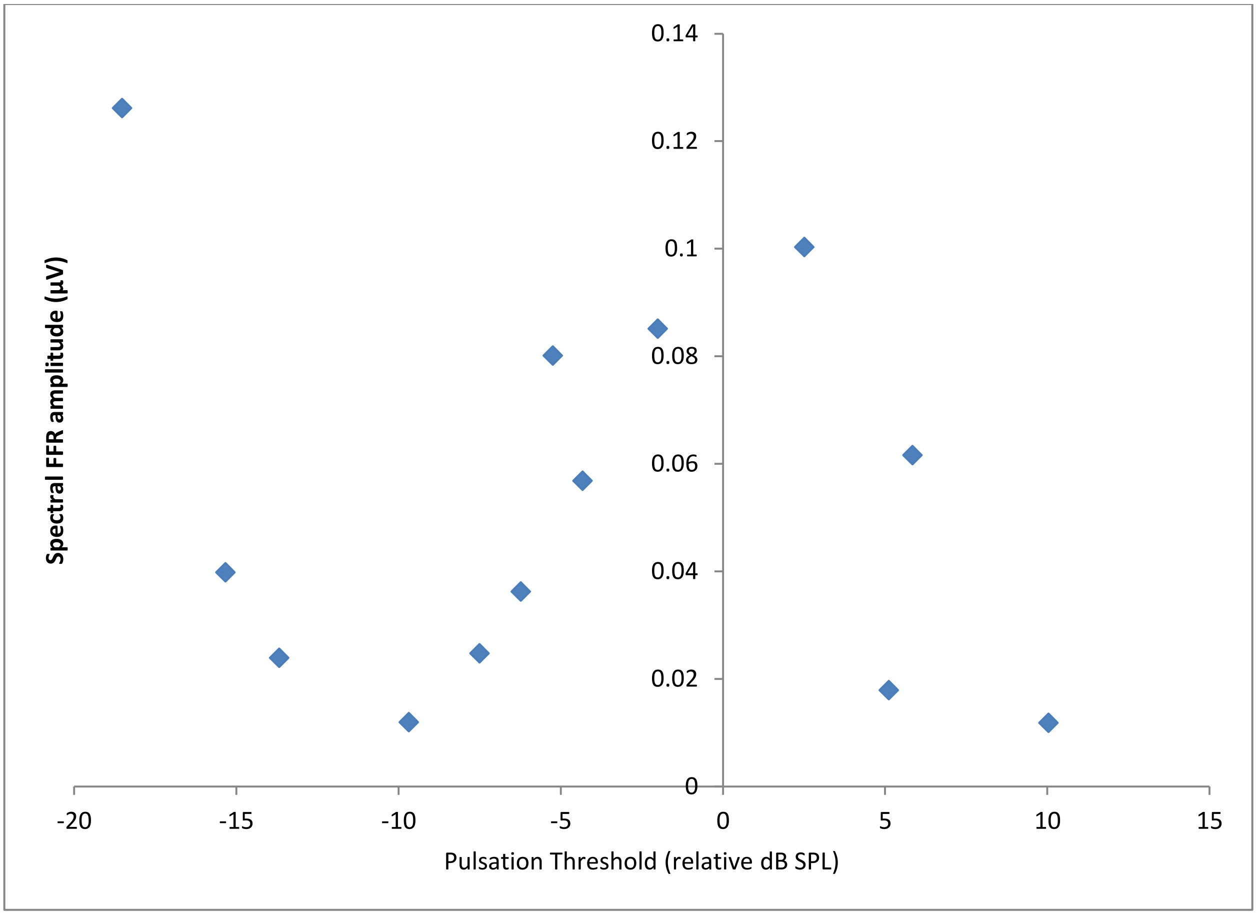
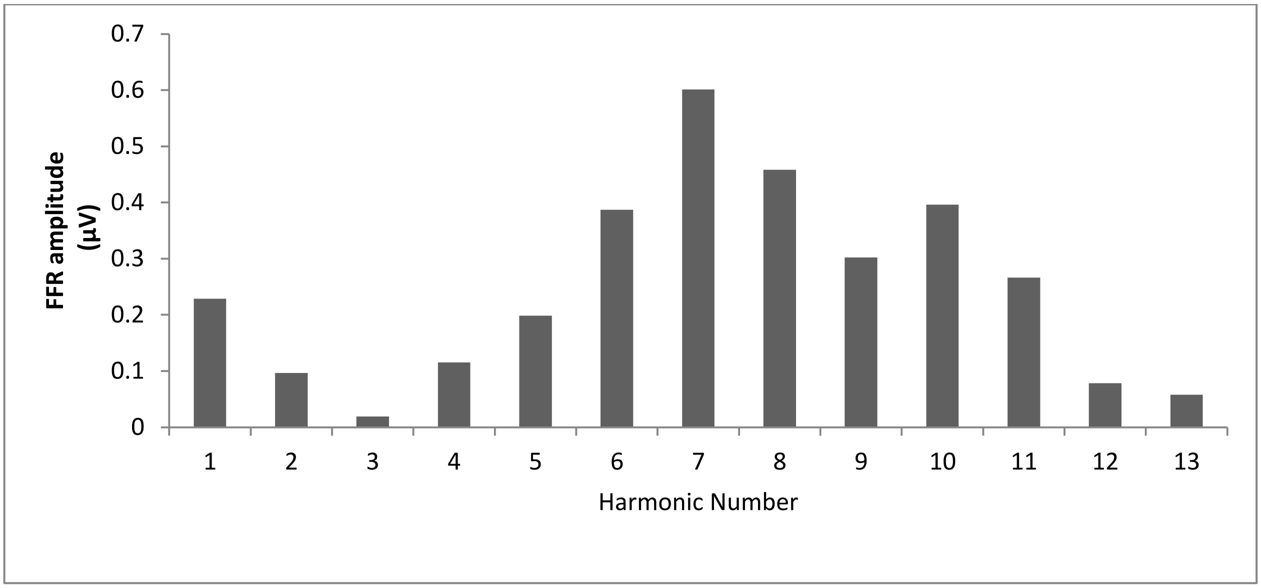


Figure 7. Grand average FFR amplitude plotted as a function of grand average pulsation threshold. The correlation between the grand average amplitude and grand average pulsation threshold across harmonics was minimal at r= -0.195. As the perceptual salience of an individual harmonic increases, the pulsation threshold will decrease to a negative number; if high perceptual salience predicted a large FFR response, a negative correlation would reflect this relationship. However, there is no evidence of such a relationship with a correlation of -0.195.



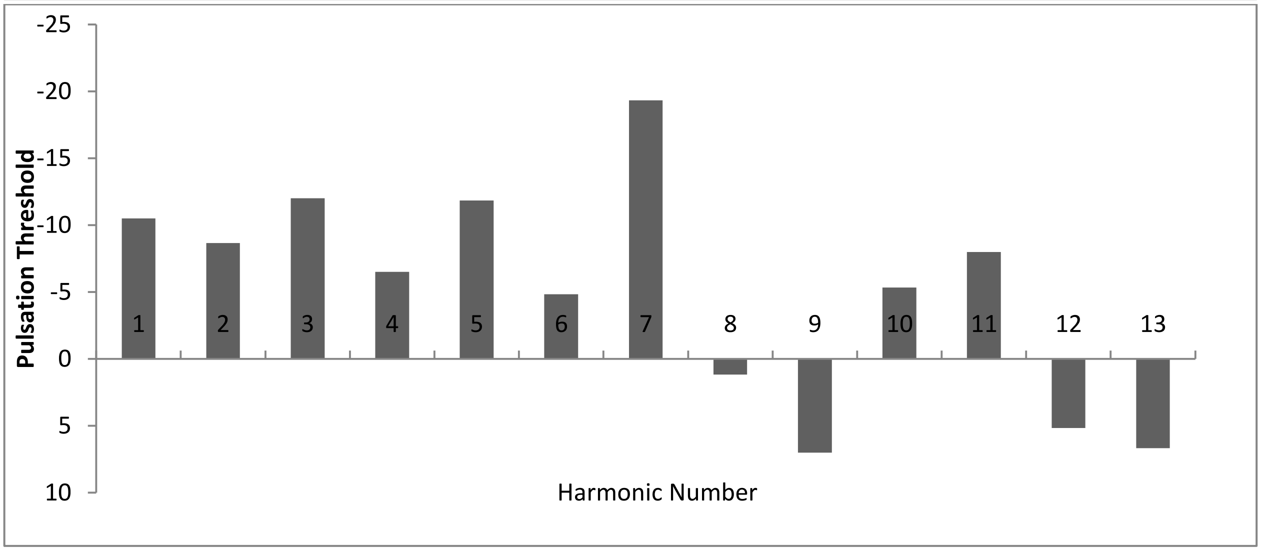
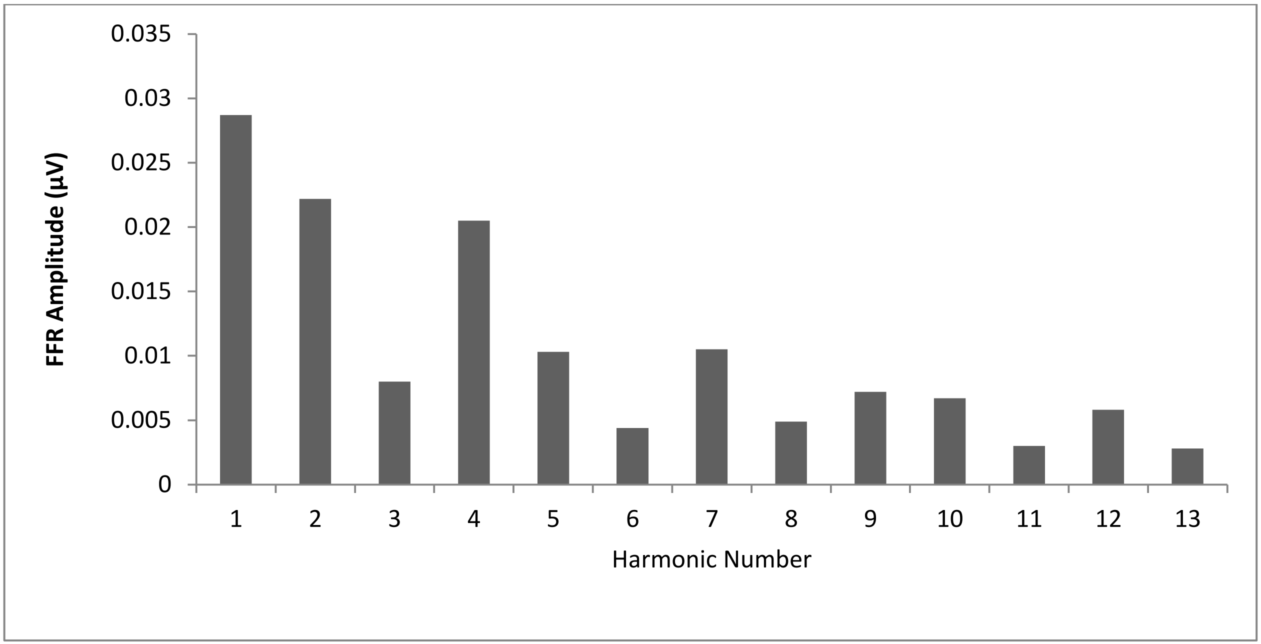


Figure 8. FFR amplitudes across harmonics for the participant with the highest average FFR (the best waveform response).

Figure 9. Pulsation thresholds across harmonics for the participant with the highest average FFR. While the highest perceptual salience value matches the highest FFR amplitude at H7, there is no other predictable relationship at the other harmonics.



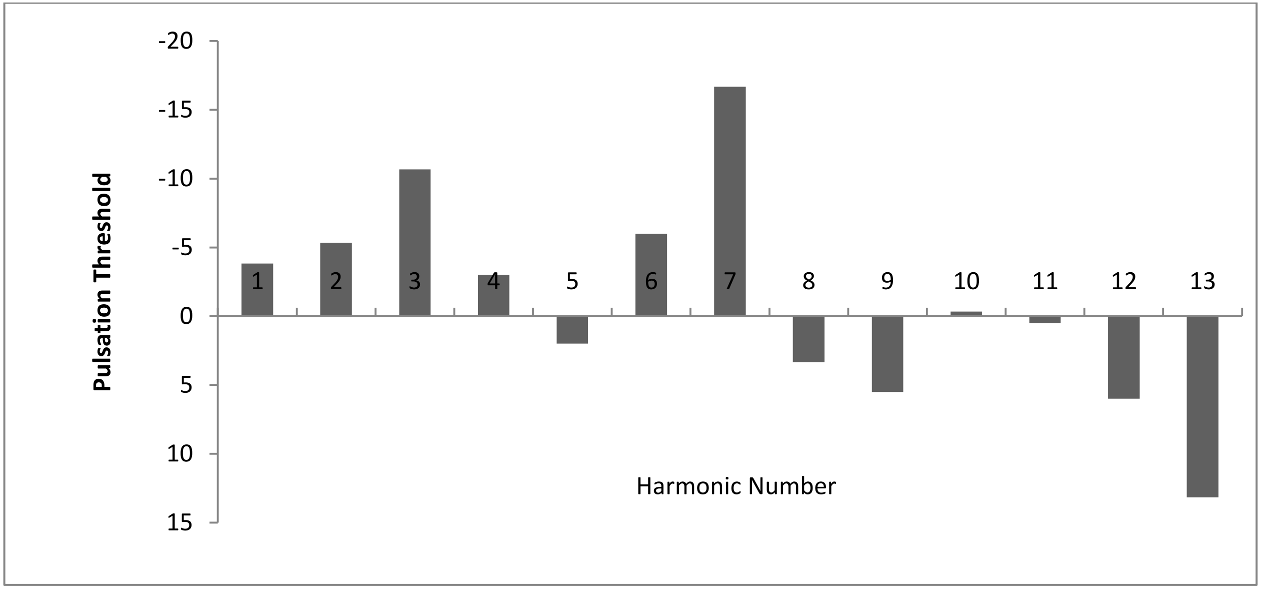


Figure 10. FFR amplitudes across harmonics for the participant with the lowest average FFR (the worst waveform response).

Figure 11. Pulsation thresholds across harmonics for the participant with the lowest average FFR. There is no other predictable relationship between perceptual salience and brainstem response at the harmonics.