Human Frequency Following Responses to Voice Pitch: Relative Contributions of

the Fundamental Frequency and Its Harmonics

A Thesis Presented to

The Honors Tutorial College

Ohio University

In Partial Fulfillment

of the Requirement for Graduation from

The Honors Tutorial College

with the degree of

Bachelor of Science in Hearing and Speech Sciences

by

Cassie E. Costilow

June 2010

ABSTRACT

The phenomenon of the "missing fundamental frequency" has shown that when the f0 is removed from a complex stimulus the pitch of the f0 is still perceived. This ability for normal hearing adults to process changes in voice pitch has been studied through psychophysical experiments. Through the use of the frequency following response (FFR), relative contributions of the f0 and its harmonics to pitch perception can be examined to determine the role of place cues and temporal cues in pitch processing in the human brainstem. The current study examined the contribution of the f0 and its harmonics in pitch processing by systematically manipulating the speech stimulus to remove component frequencies. It was hypothesized that as the f0 and part of its harmonic components were removed from the stimulus, FFRs would remain stable (in support of the temporal theory), while a response would also be identifiable when only the f0 is preserved (in support of the place theory). FFRs were recorded to seven experimental conditions including the intact, no-f0, harmonics-only and f0-only conditions. A control condition was conducted with the sound tube plugged and moved away from the participants. The results showed distinguishable FFRs in all conditions (except the control condition), with significantly larger FFR Pitch Strength in response to the harmonics-only conditions than those obtained in the f0-only condition (one-way ANOVA, p<0.001). This finding was in support of both the temporal and place theories, with temporal cues contributing more to pitch processing in the human brainstem than place cues.

ACKNOWLEDGEMENTS

I would like to thank Dr. Fuh-Cherng Jeng, my thesis advisor, for his continued support, guidance, and patience during the past two years. This thesis would truly not have been possible without his tremendous ability to explain complicated processes in manner that can be understood by everyone. Thank you for being my mentor and encouraging me throughout this process.

Secondly, I would like to thank Dr. Chao-Yang Lee, the director of studies for HTC, for his encouragement to take on this challenge, and continued support throughout the program. I would also like to thank Ms. Teresa Tyson-Drummer, Dr. Brooke Hallowell, the Hearing, Speech and Language sciences department, and Honors Tutorial College staff for their help and support throughout my academic career.

Finally I would like to thank the graduate audiology students, Brenda Dickman, JiongHu, and Daniela Stangherlin, for their encouragement and assistance.

TABLE OF CONTENTS

3

Abstract1
Acknowledgements2
List of Tables4
List of Figures
Introduction
Hypotheses9
Methods10
Participants10
Auditory Stimulus10
Recording Procedures and Experimental Design11
Data Analysis12
Results13
Stimulus Spectrograms13
Spectrograms of Typical Recording15
Objective Measures17
Results of one-way ANOVA19
Discussion
References25

Page

LIST OF TABLES

LIST OF FIGURES

Figure 1: Stimulus Spectrograms	14
Figure 2: Spectrograms of Typical Recording	16
Figure 3: Objective Measures	18

INTRODUCTION

The human brain is capable of discriminating subtle changes in voice pitch from speech signals, which is a necessary ability for the development of language. Voice pitch can carry linguistic or emotional information that is needed to understand the speaker's message. Speech signals, like other complex sounds, consist of a fundamental frequency (f0) and component frequencies that are integer multiples of the f0, known as harmonics (Ballantyne, 1990). While the f0 is known to carry vital information of the sound, the harmonics also play an important role in pitch processing, as shown by the phenomenon of the "missing fundamental frequency". This phenomenon has revealed that when the f0 is removed from a complex stimulus the pitch of the f0 is still perceived (Aiken &Picton, 2006; Dajani, Purcell, Wong, Kunov&Picton, 2005). Research using behavioral and electrophysiological methods (Aiken &Picton, 2006; Dajani et al., 2005; Zatorre, 1988) indicated that harmonics provide adequate information for pitch processing to occur in absence of the f_0 , but further exploration into the neural processes behind this phenomenon will give us a more detailed understanding of how the brain processes pitch information.

There are currently two theories that seek to explain the physiological mechanisms underlying neural encoding of complex sounds. The first, known as the "place theory" (e.g. Young & Sachs, 1979), asserts that encoding of complex sounds is determined by a location of frequency specific neurons in the auditory system that are activated starting from the basilar membrane at the cochlear level. The second theory asserts that encoding of complex sounds in the brainstem is phase-locked and based on

temporal aspects of the stimulus such as the individual harmonics and the periodicity of the waveform. Known as the "temporal theory" (Krishnan & Parkinson, 2000; Krishnan, Gandour&Cariani, 2004), studies investigating the phenomenon of the "missing f0" have provided evidence in support of this theory. When energy at the f0is removed, the brain must rely on temporal cues to encode pitch information rather than solely relying on place cues. Our study seeks to determine how large of a role, if any, each of these assumptions plays in pitch processing.

The ability for normal hearing adults to process changes in voice pitch has been studied with the use of the frequency-following response (FFR) (Aiken &Picton, 2006; Aiken &Picton, 2008;Dajani et al., 2005; Krishnan et al., 2004). The FFR is a scalp-recorded auditory evoked potential that follows the pitch contour of a complex stimulus, reflecting neural phase-locked activity (Moushegian, Rupert &Stillman, 1973). Because this measure (1) does not require the participant's response, (2) is an objective and non-invasive method, and (3) is sensitive to changes in voice pitch (Krishnan et al., 2004), it is an ideal method for exploring the phenomenon of the "missing f0." Previous studies exploring this phenomenon have studied the effects of removing the f0 from the stimulus with no investigation into the role each harmonic plays in an individual's ability to process pitch information.

Dajaniand colleagues (2005) compared FFR recordings to an intact stimulus representing the natural vowel /a/ and a token in which the energy at the f0 had been suppressed using a high-pas filter. Results of this study found evoked responses that were equally strong for the intact and missing f0 conditions. In 2006, Aiken and Picton

examined the role of the f0 in pitch processing by presenting subjects with a no-f0 and f0-only condition. FFR recordings to the natural vowel / / in the no-f0 condition revealed significant responses at the fundamental when no energy was present, while recordings for the f0-only condition revealed no significant responses. The results of the first condition provide evidence for the "missing fundamental", while the second condition suggests that harmonics play a crucial role and are necessary for pitch processing. While these findings provide a basis for understanding the role of the f0 in pitch processing, questions remain concerning the function of harmonics in this process.

The purpose of the current study was to examine the contribution of the *f*0 and each harmonic in pitch processing by systematically manipulating the speech stimulus to remove component frequencies. This includes removing f0 and harmonic frequencies separately. In an effort to evoke strong FFRs, the rising Mandarin tone "yi" was chosen as the stimulus. Krishnan and colleagues, (2004) demonstrated that a rising tone produces a stronger FFR than the level, falling-rising, or falling Mandarin tones. The results of this study will serve to help fill in the gaps of our knowledge of how the brain processes pitch information of signal processing (Ballantyne, 1990). The understanding of speech processing is applicable in daily life. By increasing our knowledge of how the brain processes sound, knowledge obtained from this study might be useful to help improve speech-processing strategies for individuals with hearing loss, such as hearing aid or cochlear implant users.

Hypotheses

The hypotheses of this study were as follows:

- (1) As frequency components are removed from the stimuli, FFR recordings will show a decreasing ability of the brain to process pitch information until no response can be recorded.
- (2) A slight response will still be present at the removal of the second harmonic.
- (3) A response will no longer be found after the removal of the 4th harmonic, due to the fact that neural phase-locking decreases substantially for frequency components \geq 1000 Hz.
- (4) Also, it is hypothesized that a response will be seen when only the fundamental frequency is preserved.

METHODS

Participants

A total of 17 adult participants were recruited from the Athens, Ohio and Ohio University communities. Participants were native speakers of English ranging in age from 19-28 years (21.75 ± 2.89 years). All participants possessed normal hearing as assessed by pure tone audiometry with hearing thresholds of 20 dB HL or better for octave frequencies between 125-8000 Hz. Participants were asked to rest comfortably while seated in a recliner. Due to the sensitive nature of the frequency-following response, only participants that were able to achieve a restful state were included in this study. If any one trial from a single recording session failed to meet the rejection rate criteria of less than 10%, all data for that participant were excluded from analysis. As a result, data from 12 participants were included for further analysis.

Auditory Stimulus

A total of seven acoustic stimuli were prepared for this experiment. A monosyllabic Mandarin Chinese syllable representing the rising lexical tone yi2 served as the intact stimulus. The syllable was recorded by a native Mandarin Chinese adult male speaker, with a sampling rate of 40000 samples/sec. The subsequent stimuli were variations of the intact stimulus, bandpass filtered to remove frequency components of the *f*0 and harmonics. Experimental conditions included the following stimuli: intact, *-f*0 (high-pass filter cutoff frequency 170Hz), *-*h2 (340Hz), *-*h4 (680Hz), *-*h6 (1020Hz), *-*h8 (1360Hz), and *+f*0 (low-pass filter cutoff frequency

170Hz). A control condition in which no sound was presented (insert ear tip was removed and plugged) was also completed to examine the amount of residual stimulus artifact. Each token had a duration of 250 ms, with a rise and fall time of 10 ms and a silent interval of 45 ms. Tokens were compensated to the root-mean-square amplitude to ensure each token contained the same amount of acoustic energy. Stimuli were presented monaurally through an ER-3A insert earphone to the right ear at a level of 70 dB SPL with a total of 2200 sweeps for each condition.

Recording Procedures and Experimental Design

Recording took place in an acoustically and electrically treated sound booth. Three surface electrodes provided a non-invasive means of recording the frequency following response. A one-channel recording montage utilized the high forehead (Fpz) as the active, non-inverting channel, the mastoid of the test ear (M2) as the reference, and the mastoid of non-test ear (M1) as the ground. The non-test ear was plugged to prevent interference from ambient noise. All electrode impedances were maintained under 3 $k\Omega$. Stimuli presentation was controlled through a custom program written in LabView (National Instruments, Austin, TX). Tokens were gradually increased from an inaudible level to the testing level. Recording began once the target stimulus presentation level was reached.Continuous data was recorded using the NeuroScan ACQUIRE 4.4 software, bandpass filtered at 0.05–3500 Hz. Data was stored on a personal computer offline analysis.

Data Analysis

All data was analyzed using MatLab (Mathworks, Natick, MA). A total of 2200 sweeps were recorded for each condition. A single sweep would be rejected if it contained voltages greater than $\pm 25 \,\mu$ V. A rejection rate of less than 10% was set for each trail, with the remaining sweeps averaged. Cross-correlation was preformed on the stimulus and recorded waveformto find the location that had the largest cross-correlation value and a 250 ms extract of the data was taken from there. Data obtained from each trial was analyzed separately. Four objective measures were used to analyze the data for each condition; *Frequency Error, Slope Error, Tracking Accuracy* and *Pitch Strength*.

Frequency Error represents the accuracy of pitch tracking during stimulus presentation. *Slope Error* indicates how well the brain follows the overall shape of the pitch contour. *Tracking Accuracy* reflects the overall accurateness of pitch encoding in the brainstem. *Pitch Strength* reflects the robustness of the response. A one-way repeated measures ANOVA was completed on the four measures to determine significance between testing conditions. Only testing conditions were included in statistical analysis. The Tukey-Kramer test, a strict post hoc test, was also conducted to determine significant conditions within each measure. A *p*-value of <0.05 was considered statistically significant.

RESULTS

Stimulus Spectrograms

Stimulusspectrograms are plotted in Figure 1. The spectrogram of the intact stimulus reveals energy bands at the f0 and multiple harmonics, with the strongest energy bands present at h2 and h3. In the stimulus spectrogram of the -f0 condition, removal of acoustic energy using a high-pass filter with a cutoff frequency at 170 Hz is seen. This pattern of removing acoustic energy below the designated cutoff frequency continues to the -h8 condition, with energy below 1360 Hz removed. The stimulus spectrogram of the remaining +f0 condition reveals an energy band only at the f0, with the removal of acoustic energy using a low-pass filter at 170 Hz.



Figure 1. Stimulus spectrograms for the seven testing conditions: intact, no fundamental frequency (-f0), removal of the 2nd harmonic (-h2), removal of the 4th harmonic (-h4), removal of the 6th harmonic (-h6), removal of the 8th harmonic (-h8), and fundamental frequency only (+f0).

Spectrograms of Typical Recordings

Spectrograms of FFR recordings revealed energy at the f0 for each of the testing conditions. Though variability in FFR spectrograms exists among subjects, the overall trend remains the same across subjects. FFR spectrograms of the intact condition displayed a typical response at the f0. When the f0 was removed from the stimulus, FFR spectrograms similar to that of the intact condition were observed. FFR spectrograms of the -h2, -h4, -h6, and -h8 conditions demonstrated a strong response to the f0 though energy at the fundamental was only present in the intact condition. FFR spectrograms of the +f0 condition demonstrated weak responses compared to the remaining testing conditions, though a response was still present. Note that spectrograms for the control condition did not show any consistent energy patterns.

An example of a typical FFR recording spectrogram is plotted in Figure 2. In the figure, clear responses are observed in the intact, -h2, -h4, and -h6 conditions. Energy at the *f*0 remains present in the -f0 and -h8 conditions, though a clear energy band is not observed. The +f0 condition reveals energy at the *f*0, though the response was weaker than the other conditions. The spectrogram of the control was free of any noticeable energy that followed the *f*0 contour of the stimulus.



Figure 2.Spectrograms of typical recording from a participant (Subject 10) for all of the testing conditions, as well as a control condition. The testing conditions included: intact, no fundamental frequency (-f0), removal of the 2nd harmonic (-h2), removal of the 4th harmonic (-h4), removal of the 6th harmonic (-h6), removal of the 8th harmonic (-h8), and fundamental frequency only (+f0).

Objective measures

Group results of the 12 participants for each of the objective measures are displayed in Figure 3. The overall trend of the *Frequency Error* remained relatively stable from the intact condition through the removal up to the 8th harmonic, all falling roughly falling between 5-9 Hz. *Frequency Error* began to rise at the +f0 only condition, with responses at the control condition showing the highest error rate between 18-19 Hz. *Slope Error* followed the same overall trend, with a rise in slope error beginning at the +f0 only condition and rising further in the control condition.

Tracking Accuracy followed a relatively stable pattern from the intact condition through the removal up to the 8th harmonic, with a decreasing accuracy of pitch encoding beginning with the +f0 condition. The control condition displayed a *Tracking Accuracy* mean value of 0.17, with no response present. The lack of response in this condition indicated a good quality of recording and appropriate elimination of stimulus artifacts. *Pitch Strength* displayed a relatively stable pattern from the intact through the removal of harmonic frequencies with a slight rise at the removal of the 6th harmonic. This was followed by a marked decrease at the +f0 condition and the control condition. Overall, the -h6 condition showed the strongest response across the eight conditions.



Figure 3.Objective Measures: Group data from the 12 participants, graphed into *Frequency Error, Slope Error, Tracking Accuracy*, and *Pitch Strength*. Conditions are as follows: intact, no fundamental frequency (-f0), removal of the 2nd harmonic (-h2), removal of the 4th harmonic (-h4), removal of the 6th harmonic (-h6), removal of the 8th harmonic (-h8), fundamental frequency only (+f0), and control (ctrl).

Results of one-way repeated measures ANOVA

Results of statistical analysis found all four objective measures to be significant with a p-value less than 0.05. Further post hoc analysis identified specific conditions within each measure that were significant from each other. For clarity, a complete list of significant conditions is displayed in Table 1. Within each of the four measures, the -h6 condition was significantly different from at least one other condition, which always included the +f0 condition. *Pitch Strength* showed the greatest significant differences between conditions with a total of five significantly different conditions, followed by *Tracking Accuracy* with three, and *Frequency Error* and *Slope Error* with one each.

DISCUSSION

The results of this study have demonstrated that both the fundamental frequency and harmonics contribute significantly to pitch processing. As expected, a robust response was observed at the f0 in the intact condition. The results of previous studies have reported the possibility of eliciting such a response to the f0 of complex stimuli in speech and non-speech contexts (e.g. Aiken et al., 2008; Krishnan & Parkinson, 2000; & Krishnan et al., 2004). The robust response to the f0 observed in the -f0 condition, was also expected due to previous research exploring the "missing f0" phenomenon (Aiken &Picton, 2006; Dajani et al., 2005). The current study was most concerned with the results of the harmonic and fo only conditions. Robust responses at the f0 were observed in FFR spectrograms for the -h2, -h4, -h6, and -h8 conditions, providing strong evidence in support of the temporal theory. While weaker than harmonic conditions, a response to the f0 was observed in the +f0 condition. There was a significant difference observed between the +f0 condition and several other testing conditions across objective measures. This condition served as a means to examine the relative contributions of the +f0 condition, and in the process lent support for both the place and temporal theories.

While this study has provided physiological evidence for both the temporal and place theories of pitch processing, it is the temporal theory that appears to hold more significance. Due to previous studies examining the phenomenon of the "missing f0", the temporal theory was already supported as a means for explaining pitch encoding in

the brainstem. The results of this study provide additional evidence for this view of pitch encoding in the human brainstem due to the lack of energy in the stimulus spectrograms that was still well represented in the FFR spectrograms.From these results, it was concluded that the brain does indeed rely on temporal aspects of the stimulus waveform, such as the periodicity, to fill in the gaps for the encoding of pitch information.

Although the temporal theory seems to emerge as the leader, the place theory is also of importance. The response at the +f0 condition indicates that there may be a population of neurons that are frequency specific, though the response is not as robust as when temporal cues are provided. The significantly lower response observed for the +f0 condition compared to the harmonic conditions provides evidence for this view. The most notable indicator was *Pitch Strength* in which the results of the +f0 condition were significantly different from the -h4, -h6, and -h8 conditions. In a 2004 study, Krishnan and colleagues studied the effects of directional pitch change on FFR recordings, and in the process discovered that all of the FFR recording spectrograms were dominated by frequencies related to the f0 and up to the first formant frequency. Because of similar findings by previous studies, such as one competed by Young and Sachs (1979), Krishnan and colleagues (2004) suggested that the strong response might indicate a single nerve population responding to the formant frequency. While this may have been true, it seems that the relative contributions of harmonics and temporal cues appear to be greater than the contributions of the f0 and place cues for pitch processing in the brainstem.

In contrast to the original hypotheses, the ability for the brain to process pitch information did not decrease as frequency components were removed from the stimulus, but instead showed a slight improvement when higher harmonics were removed. A possible explanation for these results could be in the creation of the stimuli. As frequency components were removed from the stimuli, energy was lost. As a result the tokens were scaled up to the same overall root-mean-square (rms) amplitude to ensure that all tokens contained the same amount of energy. This process resulted in the distribution of more energy among the remaining harmonics and likely produced a stronger response. However, if we did not equalizerms amplitude when we removed harmonics, acoustic tokens with a fewer number of remaining harmonics would contain a smaller overall amplitude than those with more harmonics remaining. In such a case, results would likely be confounded by the overall stimulus intensity of the acoustic tokens. That is, there is a trade off between the choice of equalizing rms amplitude across acoustic tokens rather than maintaining the original amplitude of the f0 and each of its harmonics. In our opinion, compensating rms amplitude is beneficial because it allows us to compare results across testing conditions. This approach, however, has some drawbacks. Had this compensation not taken place, it is anticipated that there would have been decreased response amplitudes at higher harmonics.

Important to note is the significant response observed for the -h6 condition. While it was originally hypothesized that a response would not be present for this condition, there was instead a robust response that was lower in *Frequency* and *Slope Error* and higher in *Tracking Accuracy* and *Pitch Strength* than any other condition. This robust response may indicate a sound intensity tradeoff taking place in the auditory system. Krishnan (2002) examined the effect of sound intensity levels on three vowels; /u/, //, and /a/. While greater responses were seen for lower frequencies at lower intensity levels, it required a greater sound intensity level for increased responses at the frequencies of h7 and h8 to be observed. In addition, in the present study the cutoff frequency of the -h6 stimulus is 1,020 Hz which falls within the frequency range most important for speech perception. The brain may have utilized the sound energy in the stimulus to produce a strong response that other tokens would need a greater intensity level to produce.

In the future, further investigation into the contributions of the fundamental frequency and its harmonics should be completed to see if language experience has an effect on the way pitch is processed. A similar study needs to be completed in native speakers of Mandarin Chinese to determine that effects, if any, of language experience. Another direction of this study should be to further investigate the role of higher harmonics to discover the point of decreased response amplitude. FFRs to the removal of the 8th harmonic were still very strong. Since FFRs diminish greatly above 2000 Hz, it would be expected that there would eventually be a lack of response.

The results of this study serve to help fill in the gaps of our knowledge of how the brain processes pitch information of complex sounds. It has also provided strong evidence for the importance of temporal cues in pitch processing. Current hearing aid and cochlear implant technology utilizes only frequency specific place cues in speech processing strategies. By continuing to gain a better understanding of how the brain processes pitch information, we hope to improve upon the current technology of hearing aids and cochlear implants to include both temporal and place cues in processing strategies. The ultimate goal is toenhance the listening experience of all hearing impaired individuals around the world.

REFERECES

- Aiken, S.J.&Picton, T.W. (2006). Envelope following responses to naturalvowels. *Audiology & Neuro-Otology*, 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.
- Ballantyne, D. (1990). *Handbook of audiological techniques*.Rushden: Butterworth-Heinemann.
- Dajani, H. R., Purcell, D., Wong, W., Kunov, H., &Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, 52(9) 1614-1518.
- Krishnan, A. (2002). Human frequency-following responses: representation of steadystate synthetic vowels. *Hearing Research*, *166*, 192-201.
- Krishnan, A., Xu, Y., Gandour, J. T., &Cariani, P. A. (2004).Human frequencyfollowing response: representation of pitch contours in Chinese tones. *HearingResearch*, 189, 1-12.
- Krishnan, A.& Parkinson, J. (2000).Representation of Tonal Sweeps.Audiology & Neuro-Otology, 5, 312-321.
- Moushegian, G., Rupert, A.L., &Stillman, R.D. (1973). Scalp-recorded early responses in man to frequencies in the speech range.*Electroencphalography* and Clinical Nuerophysiology, 35(6), 665-667.

- Young, E.D.& Sachs, M.B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *Journal of the Acoustical Society of America*, 66, 1381-1403.
- Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustic Society of America*, 84(2), 566-572.

Table 1

Results of one-way repeated measures ANOVA

	Frequency	Slope	Tracking	Pitch
	Error	Error	Accuracy	Strength
	p=0.022	p=0.029	p=0.002	p<0.001
Significant conditions	-h6:+ <i>f</i> 0 (<i>p</i> <0.05)	-h6:+f0 (p=0.034)	-f0:+f0 (p=0.022) -h6:+f0 (p<0.001) -h8:+f0 (p=.047)	-h4:+f0 (p=0.006) -h6:intact (p=0.034) -h6:-h2 (p=0.021) -h6:+f0 (p<0.001) -h8:+f0 (p=0.015)

Numbers inside parentheses are p-values indicating significance. Significance was reached if p<0.05

Terms: -*f*0 (removal of the fundamental frequency), -h2(removal of the 2^{nd} harmonic), -h4 (removal of the 4^{th} harmonic) -h6 (removal of the 6^{th} harmonic), -h8 (removal of the 8^{th} harmonic), +*f*0 (fundamental frequency only)