An Automated Response Detection Procedure for Human Frequency Following Response

Elicited by Voice Pitch

A dissertation presented to

the faculty of

the College of Health Sciences and Professions of Ohio University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy

Jiong Hu

August 2013

© 2013 Jiong Hu. All Rights Reserved.
This dissertation titled

An Automated Response Detection Procedure for Human Frequency Following Response

Elicited by Voice Pitch

by

JIONG HU

has been approved for

the School of Rehabilitation and Communication Sciences

and the College of Health Sciences and Professions by

Fuh-Cherng Jeng

Associate Professor of Communication Sciences & Disorders

Randy Leite

Dean, College of Health Sciences and Professions
Abstract

HU, JIONG, Ph.D., August 2013, Hearing Science

An Automated Response Detection Procedure for Human Frequency Following Response Elicited by Voice Pitch

Director of Dissertation: Fuh-Cherng Jeng

Frequency following response (FFR) has received a fair amount of attention in recent years. Researchers have probed different aspects of the response and yet, in most of those studies, the presence of such a response is based on subjective interpretations of examiners. Aside from a sole recent report examining two algorithms for detecting FFRs, there have been a very limited number of attempts to further develop a sound automated procedure for detecting FFRs. The purpose of this study was to (a) develop an automated procedure derived from response detection algorithms that are based upon the statistical properties of both the temporal and spectral energy distributions in the recorded waveforms, and (b) explore the effectiveness, accuracy, and efficiency of the automated procedure.
Acknowledgments

The writing of a dissertation can be a lonely and isolating experience, yet it is obviously not possible without the personal and practical support of numerous people. Thus my sincere gratitude goes to those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

I would like to sincerely thank my advisor, Dr. Fuh-Cherng Jeng, for his guidance during my PhD studies at Ohio University. It was paramount in providing a well-rounded experience consistent of my long-term career goals. I also owe my gratitude to my dissertation committee members, who offered help, guidance and encouragement in many of the stages in completing this dissertation work.

I would also like to acknowledge my peer PhD students, my lab members and friends, with whom I worked closely and puzzled over many of the same problems. They have always been there, supporting me unconditionally.

Most importantly, none of this would have been possible without the love and patience of my family. I am grateful to my wife, Dr. Jingping Hao, in particular, who has always had faith in me and supported me all these years around.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>4</td>
</tr>
<tr>
<td>List of Tables</td>
<td>7</td>
</tr>
<tr>
<td>List of Figures</td>
<td>8</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>10</td>
</tr>
<tr>
<td> Background</td>
<td>10</td>
</tr>
<tr>
<td> Specific Aims and Hypotheses</td>
<td>11</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
<td>13</td>
</tr>
<tr>
<td> History and Characteristics of Classical Frequency-Following Responses</td>
<td>13</td>
</tr>
<tr>
<td> Recent Development of Frequency-Following Responses</td>
<td>21</td>
</tr>
<tr>
<td> Response Detection Algorithms in Auditory Electrophysiology</td>
<td>32</td>
</tr>
<tr>
<td> Response Evaluation in Frequency-Following Responses</td>
<td>37</td>
</tr>
<tr>
<td>Chapter 3: Methodology</td>
<td>48</td>
</tr>
<tr>
<td> Temporal Procedure</td>
<td>48</td>
</tr>
<tr>
<td> Spectral Procedure</td>
<td>57</td>
</tr>
<tr>
<td> Comparisons with Human Judgments</td>
<td>63</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>69</td>
</tr>
<tr>
<td> Temporal Procedure</td>
<td>69</td>
</tr>
<tr>
<td> Spectral Procedure</td>
<td>76</td>
</tr>
<tr>
<td> Comparisons with Human Judgments</td>
<td>78</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>83</td>
</tr>
</tbody>
</table>
Current Study ................................................................................................................ 83
Limitations and Future Considerations ......................................................................... 96
Clinical Applications .................................................................................................... 97
References ......................................................................................................................... 99
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fundamental Frequencies (f0) of the Mandarin Chinese Syllables</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>Temporal Progression of the Pitch Variance Ratio Values</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>Pitch Variance Ratio Values as a Function of the Number of Sweeps</td>
<td>71</td>
</tr>
<tr>
<td>4</td>
<td>Relative Significance Levels as a Function of the Number of Sweeps</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>True Positive Rate of the Relative Significance Level Based Spectral Procedure</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>False Positive Rate of the Relative Significance Level Based Spectral Procedure</td>
<td>81</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Frequency-following response to 500 Hz tone burst</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Frequency-following response waveforms in response to rarefaction and</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>alternating tone bursts at 60 dB nHL</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Grand averaged frequency-following response waveforms and spectra</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>plotted as a function of stimulus intensity (dB nHL) for the vowel /u/</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Frequency-following response in response to CV /da/</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>Frequency-following response to voice pitch contours superimposed on</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>stimulus f0 contours for the four Chinese speech sounds</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Spectrograms of the stimulus and typical recordings obtained from an</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>American neonate, an American adult, a Chinese neonate, and a Chinese adult</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Fast Fourier analysis of a complex signal with time-varying features</td>
<td>37</td>
</tr>
<tr>
<td>8</td>
<td>Stimulus to response cross-correlation</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>An illustration of frequency tracking by autocorrelation</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>A typical example of fundamental frequency (f0) contours of the stimuli</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>and recordings extracted using the short-term autocorrelation algorithm and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>narrow-band spectrogram algorithm</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Flow chart of the pitch variance ratio based automated procedure for</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>detecting the presence or absence of the frequency-following response to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>voice pitch</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Flow chart of the relative significance level based automated response</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>detection procedure</td>
<td></td>
</tr>
</tbody>
</table>
Figure 13: A matrix describing the calculation of the operating characteristics .............65

Figure 14: Pitch variance ratio values obtained in the experimental and control conditions ........................................................................................................................................67

Figure 15: Temporal progression of the pitch variance ratio values along the temporal axis for all four stimulated conditions and the control condition ...........................................70

Figure 16: Pitch variance ratio values as function of the number of sweeps ...............73

Figure 17: Relative significance level values obtained in the experimental and control conditions ........................................................................................................................................76

Figure 18: Relative significance level values as function of the number of sweeps ......77

Figure 19: Receiver operating characteristics of the pitch variance ratio based temporal procedure. .................................................................................................................79

Figure 20: Receiver operating characteristics of the relative significance level based spectral procedure. .........................................................................................................82
Chapter 1: Introduction

Background

Electrophysiological measurements have an important advantage over behavioral tests in that some electrophysiological techniques do not require awareness or active feedback from the participant (Elberling & Don, 1984; Jeng, Hu, Dickman, Lin, & Lin, 2011; Jeng, Hu, Dickman, Montgomery-Reagan, Tong, Wu, & Lin, 2011; Jeng, Schnabel, Dickman, Hu, & Li, 2010; John & Picton, 2000b, 2000a; Lins & Picton, 1995a, 1995b; Lins et al., 1996; Picton, Dimitrijevic, John, & van Roon, 2001; Sininger, Cone-Wesson, Folsom, Gorga, & Vohr, 2000; van Straaten, 1999). Researchers and clinicians have used those measurements on populations that conventional behavioral tests fail or could not be reliably tested on (Choi, Purcell, & Picton, 2011; Jeng et al., 2010; Jeng, Hu, Dickman, Lin, et al., 2011; Sininger et al., 2000; van Maanen, & Stapells 2010; van Straaten, 1999). Traditionally, these electrophysiological test procedures are complex in construction and technologically heavy (Sininger, Hyde, & Don, 2002; Smits & Coppin, 2002), which consequently demands a fair amount of training to operate and to interpret the results. From a clinical point of view, if some of the procedures of those electrophysiological measurements could be standardized and automated, it could dramatically promote the efficiency and portability of the tests and more importantly, reduce the potential inconsistency in data interpretation due to subjective human judgments involved in the conventional data-interpretation process (Elberling & Don, 1984; John & Picton, 2000b; Jeng, Hu, Dickman, Lin, et al., 2011). Other types of electrophysiological measurements, such as the Auditory Brainstem Response (ABR) and
Auditory Steady-State Response (ASSR), have matured automated procedures readily available to researchers and clinicians (Picton et al., 2001; van Straaten, 1999). On the contrary, although FFR has received a considerable amount of attention in recent years (Jeng et al., 2010; Jeng, Hu, Dickman, Lin, et al., 2011; Krishnan, 1999, 2002; Krishnan, Xu, Gandour, & Cariani, 2004, 2005; Russo, Nicol, Musacchia, & Kraus, 2004; Skoe & Kraus, 2010), there has not been much of progress in the development of an automated procedure (Jeng, Hu, Dickman, Lin, et al., 2011). Thus, the overall goal of this study is to develop and validate an automated procedure for the advancement in the FFR research and the potential future application in clinical settings.

**Specific Aims and Hypotheses**

**Specific Aim 1.** To explore and evaluate whether a response detection algorithm that is based upon the Pitch Variance Ratio (PVR) index, derived from the Residue Noise Level (RNL) of the temporal property of the recorded waveforms, can detect the presence of an FFR to a satisfactory level of accuracy and efficiency.

**Hypothesis 1.** Temporal algorithm based on the PVR index can detect the presence of an FFR to a satisfactory level of accuracy and efficiency.

**Specific Aim 2.** To explore and evaluate whether response detection algorithm that is based upon the Relative Significance Level (RSL) index, derived from statistic properties of the spectral energy distribution in the recorded waveforms, can detect the presence of an FFR to a satisfactory level of accuracy and efficiency.

**Hypothesis 2.** Statistics-amended spectral response extraction algorithm can detect the presence of an FFR to a satisfactory level of accuracy and efficiency.
Specific Aim 3. To explore whether the automated procedures that are developed from the PVR-based temporal algorithm and statistics-amended spectral algorithm could provide comparable Receiver Operating Characteristic (ROC) curves, when compared with human judgment.

Hypothesis 3. The automated procedures that are developed from the PVR-based temporal algorithm and the RSL-based statistics-amended spectral algorithm could provide comparable ROC curves as compared to those obtained from human judgment.
Chapter 2: Literature Review

History and Characteristics of Classical Frequency-Following Responses

The frequency following response (also known as complex-ABR: cABR, by other researchers) is a type of brainstem neural activity that is phase-locked to the individual cycles of the stimulus waveform and/or the envelope of periodic stimuli. It reflects the aggregated (integrated over a population of brainstem neural units) neural activities. The term *frequency-following response* was first used by Worden and Marsh (1968) to describe Cochlear Microphonic (CM)-like neural components recorded directly from several brainstem nuclei. In their report, they used 20 ms tonebursts to elicit responses in cats and described the recorded waveforms as “a form of acoustically evoked electrical activity which can be observed in recordings from gross electrodes in the central auditory pathway.” They found such waveform was fundamentally different than CM in that it had a sharper onset, a latency appropriate to the place from which it was recorded (round window), an amplitude burst at the onset and a decrement of amplitude over time. A similar work by Jewett & Williston (1971) used differential recording design to elicit such response from human scalp, where they employed averaging technique to acquire waveforms with higher quality and found human evoked potential that mimics the stimulus for the first time (Jewett & Williston, 1971). Inspired by these reports, Moushegian, Rupert, & Stillman (1973) suggested that using tones of different frequencies to elicit ABR could yield frequency specific information. It was during their work (Moushegian et al., 1973) that they used short tonebursts in the speech frequency range and found waves in their recordings with inter-peak intervals that corresponded to
the periods of the tonal stimuli. Because of the difficulty in isolating FFR from CM and/or stimulus artifact in the early days, the scalp-recorded FFR did not receive considerable amount of evaluation until recent years. After researchers began to realize the potential value of using FFR as a tool to investigate the role of neural phase-locking in encoding complex stimuli and the encoding of voice pitch, there have been some renewed interests in recent years.

Figure 1. Frequency-following response to 500 Hz tone burst. Responses evoked at frequencies indicated. The sounds were 70 dB above the subject's threshold for each frequency. A: The masked neural response to a 70 dB repeating tone (500 Hz) and a continuous white noise which masked the perception of the tone (compare with E). G: Stimulus artifact for the 500 Hz sound. Vertical calibration for A-F equals 0.5 µV. Upward deflections indicate positivity at the vertex electrode. Bin width 62.5 µs, total sweep time was 64 ms. Adapted from “Scalp-Recorded Early Responses in Man to Frequencies in the Speech Range,” by G. Moushegian, A. L. Rupert, and R. D. Stillman, 1973, Electroencephalography and Clinical Neurophysiology, 35, p. 666. Copyright 2011 by Elsevier. Reprinted with permission.
Figure 2. Frequency-following response waveforms in response to rarefaction (top) and alternating (bottom) 500 Hz tone burst at 60 dB nHL. The onset and the FFR components are identified in the top waveform trace. Not the doubling in frequency of the sustained neural components. Adapted from “The Frequency Following Response and the Onset Response: Evaluation of Frequency Specificity Using a Forward-Masking Paradigm,” by A. Krishnan, and J. J. Durant, 1992, *Ear and Hearing, 13*(4), p. 230. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.

Classic FFRs were primarily elicited by tone bursts, the 500 Hz tone bursts for example, were widely used than other frequencies (see Figure 1). In that era, people used data acquisition procedures similar to that of ABRs to record FFR, the latter is smaller in amplitude and is similar to ABR in many aspects. However, since the FFR waveform essentially mimics the shape of the stimulus waveforms, researchers realized that it was very important to take precaution to make sure that stimulus artifacts are not contained in the FFR recordings (see Figure 2). Further, due to the fact that the longer tone burst stimuli used as the stimuli temporally overlap the FFR, it was also very important to
ensure there is no contamination of the CM in the FFR recordings. Several observations were made (Worden & Marsh, 1968) when FFR that was recorded in some brainstem nuclei were compared to the CM recorded on the round window. First, the latency of the two differs: Latency of the CM was usually observed at 1-2 ms, while FFR, which originated from more rostral generators, had latencies that were progressively longer. The spectral properties of these two were also different: While the FFR contains spectral energy that was more concentrated at the harmonics of the fundamental frequency, the CM’s spectrum showed less harmonic distortion and became more sinusoidal. Further, FFR disappears earlier than CM in terminal anoxia, indicating the dependence of FFR on normal metabolism of neural tissue. Finally, consistent with the peripheral auditory phase-locking ability of single units, the upper frequency limit of FFR was only approximately 5 kHz, which was different from that of CM (up to the limit of hearing).
Figure 3. Grand averaged Frequency-following response waveforms (left panel) and spectra (right panel) plotted as a function of stimulus intensity (dB nHL) for the vowel /u/. The stimulus waveform and its spectrum (with F1 and F2 harmonics identified) are at the bottom of each panel. The amplified inset in the FFR spectral data clearly shows harmonic peaks in the F2 region.

Aside from the differences between FFR and CM, early researchers had also reported several aspects of classic FFR. The first aspect was latency: Researchers (Huis in't Veld, Osterhammel, & Terkildsen, 1977) suggested that unlike ABR, the onset latency of classic human scalp-recorded FFR (response to 500 Hz tonebursts) was found to be around 5.5 to 7 ms. Also, the latencies of FFRs remained relatively stable and do not change considerably over a wide range of stimulus levels (Greenberg, Marsh, Brown, & Smith, 1987) or a wide range of masking noises (Marsh, Brown, & Smith, 1975).
Further, the effect of stimulus level was also reported (see Figure 3): similar to that of the ABR, the amplitude of FFR increases with an increase in the stimulus level up to around 65 to 75 dB sound pressure level (SPL), then asymptotes or decrease with further increment in the stimulus level (Marsh et al., 1975). FFRs could also be elicited by more complex stimuli also behave similarly, in terms of the effect stimulus level to the FFR amplitude (see Figure 4) (Krishnan, 1999, 2002; Pandya & Krishnan, 2004). For example, when a two-tone approximation of a steady-state vowel (Krishnan, 1999, 2002) was used to elicit FFR, the amplitude of the second formant in FFR increased gradually when stimulus intensity was increased from 55 dB SPL to 85 dB SPL. It is believed that the increase in the number of neural fibers that are phase-locked to the stimulus and the increase in synchronized discharged rate of auditory neurons contribute collectively to the above mentioned increase in the FFR amplitude as the stimulus level increases.

The effects of stimulus frequency on the scalp-recorded classic FFR were also reported over the years. FFR reflects primarily the result of the phase-locking mechanisms in the auditory brainstem up to the inferior colliculus (Starr & Hellerstein, 1971). Thus the frequency range which classical FFRs could be recorded is related closely to the range of the phase-locking phenomena at the brainstem level (Krishnan, 2004). A steady FFR could be found at as low as 50 Hz (Gerken, Moushegian, Stillman, & Rupert, 1975; Glaser, Suter, Dasheiff, & Goldberg, 1976; Marsh, Brown, & Smith, 1970), and the upper frequency range that has been reported could go up to 2000 Hz in toneburst elicited FFRs (Galbraith, Arbagey, Branski, Comerci, & Rector, 1995; Galbraith & Doan, 1995; Greenburg et al., 1987; Moushegian, Rupert, & Stillman, 1973).
In speech-voice elicited FFRs, some frequencies up to the second formant of an English vowel, which is approximately 800 to 2200 Hz, have been reported (Krishnan et al., 2004, 2005). Further, for a 500 Hz tone burst elicited FFR, the threshold is approximately 45 to 50 dB SPL and it increases appreciably with increasing frequency (Gerken et al., 1975; Marsh et al., 1970, 1975; Moushegian et al., 1973; Rickman, Chertoff, & Hecox, 1991). Regards the amplitudes of FFR, it has been reported that the amplitude decreases with increasing frequency for tone bursts (Glaser et al., 1976; Krishnan, 1999) as well as two tone stimuli (Krishnan, 2002). Increased numbers of harmonics in a multi-tone complex (Greenberg et al., 1987; Pandya & Krishnan, 2004; Rickman et al., 1991) and in a synthetic speech sounds (Krishnan, 1999, 2002) also result in decrease in the FFR amplitude.

Another aspect that several previous researchers evaluated about the classic FFR was the distortion product of the cochlea, i.e., the characteristics of the phase-locked activity representing the neural representation of cochlear nonlinearity (Pandya & Krishnan, 2004). For example, Rickman et al. (1991) reported that for signals with F1 frequencies at 510 Hz and 800 Hz, phase-locked neural activity at F2-F1 and 2F1-F2 was observed in the FFR spectrum, across several F2/F1 ratios (1.16-1.46). Similar results were also reported for distortion products obtained from guinea pigs for the 500 Hz F1 frequency. They concluded that the F2-F1 distortion product, which was the largest and most frequently identified among all distortion products, is essentially the phase-locked component to the envelope modulation that was recorded as the auditory stead-state
response (Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour, 2012; Krishnan, Gandour, Bidelman, & Swaminathan, 2009; Rickman et al., 1991).

Finally, over the past decades, various types of stimuli have been used to elicit scalp-recorded FFRs: tone bursts (Gerken et al., 1975; Glaser et al., 1976; Moushegian et al., 1973), two-tone distortion products (Pandya & Krishnan, 2004; Rickman et al., 1991), multi-tone complex (Gardi, Salamy, & Mendelson 1979; Krishnan, 1999, 2002), speech-like signals, such as synthesized vowels and consonants (Krishnan, 1999, 2002), two-tone approximation of vowels (Plyer & Ananthanarayan, 2001), and human voices, for example the consonant-vowel (CV) syllables: /ba/, /da/, /ga/ (Banai et al., 2009; Burns, Soll, & Vander Werff, 2009; Chandrasekaran, Hornickel, King, Warrier, Hayes, & Kraus, 2002; Cunningham, Hornickel, Skoe, Nicol, Zecker, & Kraus, 2009; Johnson, Nicol, Zecker, & Kraus, 2008a; Musacchia, Sams, Skoe, & Kraus, 2007; Nicol, Zecker, Bradlow, & Kraus, 2001; Parbery-Clark, Skoe, & Kraus, 2009; Plyer & Ananthanarayan, 2001; Russo, Nicol, Musacchia, & Kraus, 2004; Russo, Nicol, Zecker, & Kraus, 2005; Skoe, Nicol, & Kraus, 2009; Wible, Nicol, & Kraus, 2005) and Mandarin syllables with varying pitch contours (Li, & Jeng 2011; Jeng et al., 2010; Jeng, Costilow, Stangherlin, & Lin, 2011; Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011; Krishnan et al., 2004, 2005, 2009; Krishnan & Gandour 2009; Song, Skoe, Wong, & Kraus, 2008; Xu, Krishnan, & Gandour, 2006; Wong, Skoe, Russo, Dess, & Kraus, 2007).
Recent Development of Frequency-Following Responses

Speech-like and speech sound stimuli: as shown in psychoacoustic studies, the first two formats in vowels are critical cues that are needed to identify vowels (Krishnan et al., 2004, 2005). The first two formats fall well within the frequency range that scalp-recorded FFRs could be obtained. So, aside from scalp-recorded FFRs that are elicited by less complex tones, the real aspect that has grasped the interests of researchers in recent years is the speech-like and human speech sound elicited FFRs. Steady-state speech-like stimuli, such as two-tone approximations of vowels (Krishnan, 1999) and steady-state synthesized speech sounds (Krishnan, 2002) have elicited FFRs that have clear peaks representing the robust neural phase-locking to the formants, as well as the harmonics. This suggests that the phase-clocked activities from the formants and harmonics in distinct populations of brainstem auditory neurons were preserved in the FFR (Krishnan, 1999, 2002).

Interest in speech sound elicited FFRs have progressed to time-variant stimuli, such as, tonal sweeps, CVs and voice pitch. Krishnan and Parkinson (2000) first evaluated the encoding of some simple rising and falling tone sweeps (400-600 Hz and 600-400 Hz, respectively). The temporal waveforms of the FFR recordings they obtained had clear peaks that corresponded to the frequency changes, indicating that a progressive shift in the population of neurons phase-locked to the changing stimulus frequency was involved. The spectrograms (see Figure 5) also showed distinctive spectral energies along the frequency trajectory, as well as an amplitude change along the temporal progression, which suggests the view that neural phase-locking decreases with increasing frequency.
(Krishnan et al., 2004, 2005). They also found that the response amplitude of the rising tone was larger than that of a falling tone, suggesting that a rising frequency trajectory could result in greater neural synchrony than a falling trajectory. Further, Bidelman et al. (2009) used tonal sweeps that represent the approximate the duration and direction of formant transit seen in certain consonants within CVs to elicit FFR in children with specific language impairment and reported that compared to their typically developing counterparts, children with specific language impairment were shown to have degraded pitch tracking accuracy, especially in tonal sweeps with rapid frequency changes.

![Figure 4](image.png)

Figure 4. Frequency-following response in response to CV /da/. The time domain representation of a 40 ms stimulus /da/ (gray) and response (black). It includes both transient and sustained response features. The stimulus plot is scaled to match the size of the response. Hence, the microvolt bar refers only to the response. Adapted from “Auditory Brain Stem Response to Complex Sounds: A Tutorial Earing and Hearing,” by E. Skoe, and N. Kraus, 2010, *Ear and Hearing*, 189, p. 8. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.
Another important type of time-variant stimuli that are widely used in recent FFR researches is the Consonant-Vowels (CV). CVs like /ba/, /da/ and /ga/ have been used to elicit scalp-recorded FFRs in various circumstances (see Figure 4). Cunningham et al., (2001) studied the neural representation of CV sounds at the brainstem level in children with learning problems and found abnormalities in their FFR response to CV in noisy environments. The spectral pattern response FFR and the precision of key stimulus features (CV transition) were degraded. Musacchia et al. (2008) also reported that by using CV (/da/) to elicit FFR, people with prolonged musical training showed substantially larger FFR response as well as a shorter latency. The FFR enhancement also correlated with the length of the listener’s musical practice. In another report focused at the long-term, musical-training effect on subcortical neurophysiological responses, Parbery-Clark et al. (2009) noted that when FFR were elicited by a speech syllable, /da/, with presences of various background noises, musicians, who they defined as people who had at least 10 years of training and consistent practice, showed faster neural timing, enhanced representation of the harmonics in the CV and less degraded response morphology when noise were present, compared to those obtained from non-musicians. Banai et al. (2009) found that phonological decoding, measured with a test of single-non-word reading, correlated significantly with the latency of an FFR recording elicited by CV (/da/). Children with poor reading scores were found to have degraded CV-elicited FFR response timing while those good readers showed more precise performance in the temporal coding tasks. Audio-visual and auditory-only stimulation effects were reported by Musacchia et al. (2005, 2007, 2008). They used the CV /da/ as the auditory token and
videos of a male speaker articulating /da/, /du/ and /fu/ as video stimuli. They reported that the lip reading of the video delayed the FFR latency to the speech onset as well as suppressed the magnitude of the FFR response. CV elicited FFRs were also reported to vary, in latency, magnitude, and frequency-encoding accuracy when recorded in the presence of background noise (Cunningham et al., 2001; Parbery-Clark et al., 2009; Russo et al., 2004, 2005, 2008). Plyer et al. (2001) used these three CVs (/ba/, /da/, /ga/) to elicit FFR in both normal hearing and hearing impaired (mild to moderate cochlear loss) listeners. They found that clear FFR obtained from normal-hearing listeners revealed clear phase-locking activity following the second formant transition while those obtained from hearing-impaired listeners did not exhibit such following trends. They also compared the FFR data and the behavioral perception in the same hearing-impaired listeners and found a strong relationship between the reduced speech-identification scores and the degradation of the neural representation of the second formant transition.

Furthermore, in addition to manipulating stimulus parameters such as the duration of the stimulus, duration of the formant transition, and formant frequency settings, FFRs to CVs have been evaluated in a number of different populations including musicians and non-musicians (Musacchia et al., 2007, 2008; Parbery-Clark et al., 2009), where musicians with long term practices, showed significant amount of enhancement in the magnitude of FFR and shortened response latency, when FFR were elicited by CVs. Children with dyslexia, specific language impairment (SLI), and autism spectrum disorders (ASD) were also evaluated in the CV elicited FFRs (Banai et al., 2005, 2009; Banai & Kraus 2008; Chandrasekaran et al., 2009; Cunningham et al., 2001; Hornickel et al., 2009; Russo et
al., 2009). These researchers found that children with developmental dyslexia exhibited impairment in the ability to modify representation in predictable context (Banai et al., 2009; Hornickel et al., 2009). Specifically, their FFR amplitudes were significantly smaller than their normal developing counterparts, when the FFR were elicited by temporally variable CVs (/bu/-/ba/-/da/-/ba/) (Banai & Kraus 2008). Similarly, when CV elicited FFR were recorded in children with ASD, the latency (representing the neural synchrony), frequency encoding (representing phase locking) and the magnitude of FFR in quiet and in the presence of background noise, all degraded when compared with normal development children at the same age. This suggests that children with ASD may suffer from abnormalities in the brainstem processing of speech that contribute to their language impairment. Other scenarios such as the effect of short-term and long-term auditory training (Russo et al., 2008; Song et al., 2008) and the developmental characteristics across the lifespan (Anderson & Kraus, 2010; Burns et al., 2009; Johnson et al., 2008b) have also been reported.

The last time-variant type of commonly seen FFR token is voice pitch. Voice pitch is an auditory perception that reflects the psychological representation of fundamental frequency (\(f0\)) of speech signal and carries important information for the perception of speech and music. The accurate encoding of voice pitch and its temporal variation is essential and critical for listeners to perceive lexical information and prosodic cues in speech signals, as well as to appreciate melodies of music. Voice-pitch-elicited and scalp-recorded FFRs have been reported in recent years with results indicating clear and faithful correlation with behavioral tests of pitch perception. Greenburg et al. (1987)
used several low-pitch-evoking complex sounds (synthesized complex sound that consists of several frequencies individually or combined: 244 Hz, 366 Hz and 488 Hz) to elicit scalp-recorded human FFR and found that the phase-locking activity underlying FFR did carry pitch-relevant information. For example, in their study, the FFR to a complex sound with missing fundamental and to a pure tone of the same frequency (\(f_0\) of the complex sound) had similar spectrum. Further, they found that FFR spectral peak (the missing fundamental) was not the result of neural phase-locking to the stimulus envelop and did not change with the alternation in their modulation depth. Similar result was also reported by Jeng, Costilow, et al. (2011), where the fundamental frequency (\(f_0\)) is removed from a complex stimulus, the pitch of the \(f_0\) is still perceived by the listener, reflected by an FFR elicited by a Mandarin syllable /yi/.

**Figure 5.** Frequency-following response to voice pitch contours (solid lines) superimposed on stimulus \(f_0\) contours (broken lines) for the four Chinese speech sounds (yi\(^1\): flat; yi\(^2\): rising; yi\(^3\): dipping; and yi\(^4\): falling). Pitch was extracted using a short-term autocorrelation algorithm (Boersma, 1993) on multiple frames of the signal utilizing a Hanning window of effective length equal to 0.04 s. Adapted from “Human Frequency Following Responses: Representation of Pitch Contours of Chinese Tones,” by A. Krishnan, Y. Xu, and J Gandpour, 2004, *Hearing Research, 189*, p. 8. Copyright 2011 by Elsevier. Reprinted with permission.
Among other human voice pitches, the Mandarin Chinese syllables (see Figure 5), which contain varying $f_0$ contours and carry lexical meaning have been used widely in the FFR research (Jeng et al., 2010; Jeng, Hu, Dickman, Lin, et al., 2011; Krishnan, 1999, 2002; Krishnan & Gandour 2009; Krishnan, Xu, Gandour, & Cariani, 2004, 2005; Li & Jeng 2011; Song et al. 2008; Wong et al. 2007; Xu et al. 2006). Recent studies have shown that neurons in the auditory brainstem could be affected by different factors, in terms of the ability to encode low frequency voice pitch, as reflected by FFRs. Krishnan and his colleagues (2004) first used a set of four Mandarin Chinese syllables, contrasting the four lexical tones (yi$^1$, ‘clothing’, yi$^2$, ‘aunt’, yi$^3$, ‘chair’ and yi$^4$, ‘easy’), to elicit scalp recorded human FFR in Mandarin Chinese speakers. They utilized autocorrelation-based pitch extraction method to measure the outcome of the response and reported that FFR indeed preserve pitch-relevant information of all stimuli and the phase-locked inter-peak intervals followed the trajectory of $f_0$ very closely. The usage of Mandarin Chinese syllable began to see a rapid increase in relevant literatures. In 2005, the same group (Krishnan et al., 2005) carried out similar studies to show the contrastive response in tonal language speakers and non-tonal language speakers and found that, when FFR were elicited using Mandarin Chinese syllables, the Chinese group exhibited stronger pitch representation and smoother pitch tracking that the English group. In a study that aimed at the same participant comparison, Krishnan et al. (2005) used sets of iterated rippled noise to mimic that Mandarin syllables and to elicit FFR in those two participant groups. They reported that the Chinese group showed smoother pitch tracking that the English group, as well as some stronger representation of multiple pitch-relevant harmonics. The
effect of long-term music training (Bidelman et al., 2009; Wong et al., 2007) and short-
term auditory training (Song et al., 2008) both have considerable levels of influence on
the result of voice elicited FFRs. For example, Wong et al. (2007) measured FFR
responses to linguistic pitch patterns (Mandarin Chinese syllable /ma/) in ten amateur
musicians and ten nonmusicians who had no previous exposure to a tone language and
found that musicians had more robust and faithful encoding of the pitch information,
reflected by their higher amplitude of the FFR waveform and enhanced pitch-tracking
accuracy. Bidelman et al. (2009) further extended this scope to the comparison between
native Mandarin Chinese speakers, English-speaking musicians and English-speaking
nonmusicians. They used both iterated rippled noises and Mandarin Chinese syllables to
elicit FFR and found that the pitch-tracking accuracy was higher in the Chinese and
musicians than in the nonmusicians, and that pitch strength was more robust in musicians
than in nonmusicians. With regard to short-term linguistic training, Song et al. (2008)
reported that when native English-speaking adults were exposed to English pseudowords,
which were based on the /f0 of Mandarin Chinese syllable /mi/, their FFR response to the
same Mandarin Chinese syllables revealed increased pitch tracking accuracy, particularly
with the dipping contour.
Figure 6. Spectrograms of the stimulus (Mandarin Chinese syllable, yi², left column) and typical recordings obtained from an American neonate, an American adult (middle column), a Chinese neonate, and a Chinese adult (right column). A gray gradient scale on the right of the spectrograms indicates the spectral amplitudes in nV for the recordings obtained from neonate and adult participants. The spectrograms of the stimulus are plotted on a normalized scale ranging from 0 to 1. All spectrograms were obtained using a Hanning window of 50 ms in length; overlap 47.5 ms in length; and a frequency resolution of 1 Hz. Adapted from “Cross-Linguistic Comparison of Frequency-Following Responses to Voice Pitch in American and Chinese Neonates and Adults,” by F.-C. Jeng, J. Hu, B. M. Dickman, K. Montgomery-Reagan, M. Tong, G. Wu, and C.-D. Lin, 2011, *Ear and Hearing*, 32(6), p. 706. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.

Further on this topic, Children with ASD (Russo et al., 2008), developmental dyslexia (Chandrasekaran et al., 2009) and SLI (Banai et al., 2009) have shown to have degraded voice-pitch tracking accuracy and decreased FFR amplitudes. Specifically, when compared with typically developing children, those with ASD showed deficient pitch tracking ability, reflected by increased frequency and slope errors and reduced phase locking to f0 and the second harmonics. Chandrasekaran et al. (2009) reported that
typically developing children show enhanced brainstem representation of features related to voice pitch elicited FFR, while children with developmental dyslexia exhibit impairment in such ability. Other aspects in the Mandarin Syllable-elicited FFRs have been reported, including the effects of background noise (Li & Jeng, 2011), using genuine (curvilinear; Krishnan et al., 2005) and modified (linear; Xu et al., 2006) Mandarin pitch contours under speech and non-speech conditions (i.e., Mandarin and musical pitch contours embedded in iterated rippled noise) (Bidelman et al., 2009; Krishnan et al., 2009; Swaminathan et al., 2008). Li & Jeng (2011) reported that in normal hearing adults, with systematically manipulation of the signal-to-noise ratio (SNR) across stimulus intensities, FFR elicited by a Mandarin Chinese syllable remained relatively stable until SNR was degraded to 0 dB or lower. Different from using the genuine Mandarin Chinese syllables, which usually have \( f_0 \) that are curvilinear, Xu et al. (2006) used synthesized Mandarin monosyllables with linear rising and falling \( f_0 \) contours to elicit FFR in normal hearing native Mandarin Chinese speakers and English speakers. They reported that no cross-language difference in pitch strength or accuracy were observed for either tone, indicating that stimuli with linear rising/falling contours elicit homogeneous pitch representations at brainstem regardless of language experience. Finally, there has been some increased interest in the developmental trajectory of human scalp-recorded FFR to voice pitch in recent years (see Figure 6). Studies have reported the characteristics of the FFR to voice pitch in newborn infants of different parental language backgrounds (Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011). When Mandarin Chinese syllables were used to elicit FFR in neonates whose parents speak
either Mandarin Chinese or English, the linguistic background plays an important role in some of the measurement index. For example, they reported that pitch strength were significantly different in the two neonate groups. Also, they reported that the same index was found to have significant difference between the Chinese neonates and Chinese adults, indicating a possible fit for the linguistic experience model in language developmental theory. They also reported the characteristics of voice pitch elicited FFR in maturing infants aged several months (Jeng et al., 2010). They found that infants at ages of several months showed slightly larger pitch strength to that of adults and some of them showed clear progression of the ability to track and process pitch information in the stimulus. Normal-hearing children of aged 3 to 4 years showed delayed and less synchronous onset and sustained neural response activity when speech FFR were elicited, compared with 5 to 12 years old children (Johnson et al., 2008a).

These studies have demonstrated voice pitch elicited FFR exhibits some early maturation of processing in neonates, infants and children. Those results also suggest that the human auditory system exhibits developmental plasticity, in both frequency and time domains. These findings could be used to help identify neonates, infants and children that are at risk of delays in voice-pitch perception or disorders like SLI or ASD, as mentioned previously. They could also be utilized to provide new directions for preventive and therapeutic interventions for patients of all ages with central auditory processing deficits, hearing loss, and other types of communication disorders. Such direction leads to the merit of this study, which is to make an effort to make an automated FFR procedure that could be used in the future to fulfill such potential.
Response Detection Algorithms in Auditory Electrophysiology

Some of the electrophysiological measurements, such as ABRs and ASSRs, have already evolved to a level that permits their automated version to be widely used in clinical settings (Choi, Purcell, & John, 2011; Elberling & Don, 1984; Norton, Gorga, Widen, Folsom, & Sininger, 2000; Oudesluys-Murphy & Harlaar, 1997; Sininger, Cone-Wesson, Folsom, Gorga, & Vohr, 2000). For example, the advancements in the automated procedures of ABR have brought us very effective and efficient audiologic assessment tools. The algorithm underlying the Automatic ABR (AABR) can be traced back to the 1980s. Based on the basic assumption that the brainstem response is deterministic and the background noise is random, researchers (Delgado & Özdamar, 2004; Don, Elberling, & Waring, 1984; Elberling & Don, 1984; Sininger, 1993; Smit & Coppin, 2002) developed statistics-based algorithms to detect the presence of an ABR. Their approach relies on the quality of the ABR recordings, which includes the use of the background noise level and the overall level of the response by calculating an $F_{sp}$ value for every block of 256 sweeps. The $F_{sp}$ is the derived from the quotient of $\text{VAR}(S)$ and $\text{VAR}(SP)$, where $\text{VAR}(S)$ is the variance of digitized amplitude values across and appropriate time window of the averaged response, which includes the evoked potential (response) and the averaged background noise, and $\text{VAR}(SP)$ is the variance of the digitized voltage of a single point in the array, calculated repetitively over a block of 256 sweeps and is used to estimate the noise level in the recording. When $\text{VAR}(S)$, which represents the $\text{RMS}^2$ of the evoked potential and the averaged noise, is divided by $\text{VAR}(SP)$, which represents the $\text{RMS}^2$ of the averaged background noise, the value, $F_{sp}$,
is closely related (Elberling & Don, 1984) to the SNR$^2$ in the response and can be used to determine the strength of the neural potential, thus the quality of the response. When no neural potential is evoked, the value of $F_{sp}$ would be expected to be close to 1, as no signal+ noise divided by noise when signal is close to zero. When the evoked potential is expected to exist, as the test progresses, more sweeps will be averaged and result in reduced level of noise and an increasing value of $F_{sp}$ as the overall amplitude of the response is progressively larger compared to the background noise. Since $F_{sp}$ is calculated by dividing the two variances, the distribution of its value should have a known F distribution, with (conservatively) degrees of freedom of 5 and 256 (Elberling & Don, 1984). With a table of F distribution, the examiner can determine the existence of an ABR when $F_{sp}$ value of 3.1 (Elberling & Don, 1984), which brings in a confidence level of 99% (1% false positive rate, $\alpha$) that there is a true response. If a lower $\alpha$ criteria is selected, the number of sweeps that is needed to reach a critical $F_{sp}$ value could reduce to a level that is more suitable for circumstances that is more time-sensitive, such as in the clinical settings.

An automated procedure could be developed based on the algorithm mentioned above: as long as the examiner determines a desired level of confidence, a pre-written computer program script could carry out the rest of the test: from the delivering the sound token, collecting electrophysiological activities, to calculating the value of $F_{sp}$ and terminate the test when the desired value is reached. It provides an effective and efficient auditory evaluation tool for researchers and clinicians by eliminating subjective human judgments, reducing the amount of training needed while at the same time providing
excellent test sensitivity and specificity (Siningger et al., 2000). Several automated ABR devices that are based on algorithms like the one mentioned above have been widely used in recent decades. Some of them use algorithms and procedures that are specifically tuned toward the use of newborn hearing screening. For example, the ALGO (Natus Medical, San Carlos, CA) system, which is based on the very algorithm that Elberling and Don described in 1984, has been widely accepted as an effective tool for newborn hearing screening (Jacobson et al., 1990; Siningger et al., 2000). It uses transient click stimuli at a level of 35 dB nHL and presents them mono-aurally at a rate of 37 clicks per second. The clicks have an acoustic spectrum flat from 750 to 5000 Hz. After the ambient noises and myogenic activities have been eliminated by preset artifact rejection settings, the algorithm uses a pre-determined template to measure the ongoing EEG for the presence or absence of the ABR. After a preset satisfaction criterion (Jacobson et al., 1990) has been reached, the system stops the data collection and display a "pass" for the testing ear.

Automated electrophysiological measurements and relative devices, such as ALGO, brings to the researchers and clinicians tools with advantages of being safe, simple to operate, and can be used by personnel who have undergone limited audiological training. Such promising features drive people looking for appropriate automated procedures in other electrophysiological measurements, such as the ASSR (John & Picton, 2000b, 2000a; Özdamar & Bohórquez, 2006; Picton et al., 1987, 2001, 2005; Valdés, Pérez-Abalo, Martín, Savio, & Sierra, 1997).

The ASSR was previously known as the "40-Hz auditory potential" because early discoveries were focused primarily on the 40 Hz modulation frequency, and was
considered to be generated from two groups of neurons: one in the auditory cortices on the top of the temporal lobe and the other in the brainstem (John & Picton, 2000b). The ASSR can be elicited by using various types of stimuli, such as sinusoidal-amplitude-modulated (SAM) tones and frequency-modulated (FM) tones, and the responses reflect the spectral energies primarily at the modulation frequency. Due to the significant amount of spectral energy at the modulation frequency in the ASSR, it makes it ideal to analyze and present the response in the frequency domain. A claimed (John & Picton, 2000b; Lins et al., 1996; Picton et al., 2001) advantage of ASSR over some other types of electrophysiological tests is the frequency specificity in its response and it is based on such frequency specificity that people have developed some automated response detection algorithms and procedures. For example, the AUDERA ERA system (Nicolet Biomedical of Viasys Healthcare, Madison, WI) employs an algorithm that uses vectors to represent in the response: the radius represents the phase angle and the magnitude represents the spectral energy, and each recording sweep resulted from the stimulation gives one vector on the circular display. The underlying philosophy is that for non-physiological activities, such as noises, the pattern of the vector distribution would be random for both the radial angles of the vectors and their magnitude. However, if the participant does hear the token and can encode the spectral information embedded in the stimulus accurately, the resulting vectors should have a certain pattern of distribution on the circular display: the response is elicited by the same frequency so that the radial angles should be clustered closely. As more and more sweeps are accounted for, there should be some large amount of energy clustered within that small sector of vector
distribution. Aside from a plot for the vector distribution, another method to demonstrate the procedure is to use a vector representing the averaged recording. Such a vector will have a radial angle representing the phase angle and a circle at the tip whose circumference represents the randomness in the averaged signal: the more noise in the signal, the larger the circle will be. Since it is based on the statistical distribution, when the circle intersects with the origin on the coordinate, the averaged signal would be considered statistically insignificant.

Another automated system used to detect an ASSR was developed by Terence Picton and Sasha John in Toronto University. Their (MASTER) system simultaneously generates multiple amplitude-modulated (AM) and/or a frequency-modulated (FM) stimulus, acquires electrophysiological responses, displays these responses in the frequency domain, and determines whether or not the responses are significantly larger than background noise (John & Picton, 2000b, 2000a; Picton et al., 2001, 2005). The system collects EEG waveforms, applies artifact rejection protocol and utilizes a fast Fourier transform (FFT) on averaged sweeps. The FFT converts the amplitude-time waveform into the spectral domain. For each frequency represented in the FFT spectrum, the X-Y (real and imaginary) coordinates are then transformed into a vector with an amplitude and phase. The amplitude is the length of the vector \( \sqrt{X^2 + Y^2} \) and the phase is the rotation of the vector in relation to the \( X \) axis as calculated by \( \arctan(Y/X) \). An \( F \)-test will then be applied onto those vectors by using the following equation to calculate the \( F \)-ratio:

\[
\frac{(x_1^2 + y_1^2)/2}{(\sum_{i=1}^{J} x_i^2 + \sum_{i=1}^{J} y_i^2)/(2J)}
\]
where \(x\) is the sine term of each polar vector and \(y\) is the cosine term, \(s\) denotes those of the signal and \(J\) is the number of vectors which are used in the summation. When the calculated F-ratio of a response at a particular modulation frequency is significant at a predetermined \(p\) level, the subject is considered to have heard the carrier frequency. (John & Picton, 2000b, 2000a; Picton et al., 2001, 2005)

**Figure 7.** Fast Fourier analysis of a complex signal with time-varying features. This response was evoked by a 40-ms /da/ sound, comprising an onset stop-burst followed by a consonant-vowel formant transition period. A frequency-domain representation of the frequency following response was generated using the fast Fourier transform (FFT). As a measure of phase locking, spectral amplitudes are calculated over a range of frequencies corresponding to the \(f_0\) (103 to 125 Hz) and the first formant (F1; 220 to 720 Hz). The noise floor is plotted in gray. Adapted from “Auditory Brain Stem Response to Complex Sounds: A Tutorial Earing and Hearing,” by E. Skoe, and N. Kraus, 2010, *Ear and Hearing, 189*, p. 9. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.

**Response Evaluation in Frequency-Following Responses**

Response evaluation in the FFR has also evolved in the recent years. For more classical FFRs, i.e., FFR elicited by tone bursts and/or steady-state complex tones,
response characteristics, such as amplitude, latency and phase information, were more commonly used (Greenburg et al., 1987; Krishnan, 1999, 2002; Pandya & Krishnan, 2004;). For example, in transient responses, people look at latencies and amplitudes for each peak, the inter-peak amplitude, duration, slope, and area (Russo et al., 2004). There are also some peak-picking algorithms that could help people to identify the maxima and/or minima of the peaks, as well as techniques such as wavelet de-noising and high-pass filtering (Hornickel et al., 2009) to help find the low-amplitude peaks. Some of these indices were measured from the magnitude spectrum obtained using Fourier analysis of the temporal domain response. Fourier analysis breaks a complex waveform consisting of many frequency components into a set of sine waves. The magnitude of each sine wave corresponds to the amount of energy contained in the complex waveform at that frequency. In FFR research, the fast Fourier transform (FFT) is the most common algorithm for performing spectral analysis (see Figure 7). For example, by performing an FFR on the prestimulus time window, the spectral noise floor can be estimated and used to calculate spectral SNRs (Russo et al., 2004).
Figure 8. Stimulus to response cross-correlation. Cross-correlation is used to compare the timing and morphology of two signals (A). The response (black, bottom) to a 170 ms /da/ (gray, top) is compared with a low-pass filtered version (dark gray, middle) of the evoking stimulus. The stimulus consists of an onset stop burst, consonant-vowel formant transition, and a steady-state (i.e., unchanging) vowel. (B) This plot represents the degree to which the low-pass stimulus and response are correlated as a function of the time shift. The maximal correlation is reached at an 8.5 ms time displacement. (C) Running-window analyses can be used to visualize and quantify the similarity of two signals across time. In this example, when the same low-pass stimulus and response are compared in this manner (40 ms windows), the two signals are more similar during the steady-state region. Adapted from “Auditory Brain Stem Response to Complex Sounds: A Tutorial Earing and Hearing,” by E. Skoe, and N. Kraus, 2010, *Ear and Hearing*, 189, p. 9. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.
For FFR elicited by time-variant stimuli, such as tonal sweep, two-tone approximation of vowels, CVs and/or voice pitch, the response evaluation processes would take some more complex approach, as the FFR recordings should not only reflect the magnitude of the response in the temporal domain, but more importantly, it should demonstrate that whether the phase-locking activity would reflect the time-varying features in the stimulus. A commonly used technique is the static or sliding window analysis, including those used in calculating root-mean-square (RMS) amplitude, cross-correlation (see Figure 8) and Fourier analysis. For example, when calculating the RMS amplitude, which represents the magnitude of neural activation over a given time period (Russo et al. 2004), data in each time window would be squared, the mean of the squared values obtained and then the square root of the means would be determined. The quotient of the response RMS amplitude would represent the signal and the prestimulus baseline RMS amplitude would act as the noise and thus a signal-to-noise ratio (SNR) can be determined. Different from the RMS measurement, which takes a static window analysis, the cross-correlation takes a sliding window approach. It is a useful tool for comparing the overall morphology and timing of two signals (Krishnan 2002; Krishnan et al., 2005). One useful application of cross-correlation in FFR studies is to objectively determine the onset of the response by correlating the stimulus and the response: the time point that yields the largest correlation coefficient will be considered the onset of the response (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011). Another commonly seen usage is correlate two responses to determine how much degrade has occurred when stimulus were presented with noise background (Russo et al.,
Aside from RMS amplitude calculation and cross-correlations, sliding window technique is also used in data visualization methods that plot three-dimensional graphs of the FFR signal, such as spectrograms and correlograms. A spectrogram has abscissa representing time, ordinate representing frequency, and the third dimension represents the spectral density energy at a given time-frequency point (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011). It is usually used to demonstrate the temporal progression of the spectral energy in the FFR recording. A correlogram has the abscissa representing time, ordinate representing the time lag between the two correlated signals, and the third dimension represents the frequency channels. It is commonly utilized to look for the periodicity embedded in the FFR recordings (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011; Krishnan, 2002; Krishnan et al., 2005). Other applications such as autocorrelation, which examines the similarity of a signal to itself to find the period of the \( f_0 \), also utilizes the sliding window technique and will be discussed in the following paragraph.
Figure 9. An illustration of frequency tracking by autocorrelation. By cross-correlating a response waveform with itself, the time interval between peaks can be determined. The frequency of the $f_0$ and other periodic aspects of the response, including the temporal envelope (Krishnan et al. 2004; Lee et al. 2009), can be derived from an autocorrelogram. This technique can also be used to calculate the strength of phase locking to these features. (A) In this example, the response to a syllable /mi/ with a dipping $f_0$ contour (Mandarin Tone 3; black line in B) is plotted. (B) By applying the autocorrelation technique on 40 ms sliding windows, a frequency contour can be tracked over time. Colors represent the strength of correlation; white is highest. (C and D) An illustration of cross-correlation performed on a single time window (100 to 140 ms; demarcated in A). When a copy of this window is shifted by 10.45 ms, the first peak of the copy lines up with the second peak of the original (C). A correlogram (D) represents the degree of correlation as a function of the time shift. The highest correlation occurs at 10.45 ms; thus, the fundamental periodicity of this window is $1/10.45$ ms or 96 Hz. The strength of the correlation at 10.45 ms is 0.98, indicating strong phase locking to 96 Hz in this time window. Adapted from “Auditory Brain Stem Response to Complex Sounds: A Tutorial Earing and Hearing,” by E. Skoe, and N. Kraus, 2010, Ear and Hearing, 189, p. 10. Copyright 2011 by Lippincott Williams & Wilkins. Reprinted with permission.
Several response evaluation algorithms have been used in recent reports. To detect the periodicity information embedded in the FFR recordings, a widely used pitch-relevant information extraction method was the short-term autocorrelation algorithm (see Figure 9). Adopted and expanded from the kernel idea reported by Boersma et al. (1973), the algorithm (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011; Krishnan et al., 2004, 2005; Wong et al., 2007) decomposes the FFR recording into successive windowed segments. Autocorrelation was conducted on each of these windowed segments. The time shift ($\tau_{\text{max}}$) which yielded the maximum autocorrelation value between certain values, depending on the $f_0$ frequency (usually 5 to 13 ms), would be identified. This range of time shifts corresponded to the fundamental frequency ($f_0$) contours of the stimuli (Mandarin Chinese syllable, yi2) and recordings extracted using the short-term autocorrelation algorithm (a) and narrow-band spectrogram algorithm (b). Adapted from “Evaluation of Two Algorithms for Detecting Human Frequency-Following Responses to Voice Pitch,” by F.-C. Jeng, J. Hu, B.M. Dickman, C.-Y. Lin, and C.-D. Lin, 2011, *International Journal of Audiology*, 50(1), p. 15. Copyright 2011 by the International Society of Audiology. Reprinted with permission.
pitch period of the stimulus $f_0$ contours (i.e., 75 – 200 Hz). The fundamental frequency of each windowed segment was calculated as $1/\tau_{\text{max}}$ (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011) and same calculation would be carried upon each windowed segment so that the resulting data sequence would be the representation of the $f_0$ embedded in the data. The magnitude of the normalized autocorrelation peak, expressed as harmonic/noise ratio ranging from 0 to 1, was also used by some researchers to represent the strength of the FFR response, denoted pitch strength (Krishnan, 2002, Krishnan et al., 2004). Another pitch information extraction method was introduced by Jeng, Hu, Dickman, Lin, et al., (2011), which used the information from the spectral-domain (see Figure 10). The “spectral-density-maxima” algorithm utilizes the spectral energy density distribution of the FFR and searches for the maximal spectral density in a predetermined frequency range in every windowed segments. In a predefined frequency range (Jeng, Hu, Dickman, Lin, et al., 2011), the algorithm searches for the frequency point that has the maxim spectral energy. This frequency point would be marked as the corresponding $f_0$ for that windowed segment and such. The algorithm then progress along the temporal axis for the rest of the data and that the resulting frequency point sequence would be considered the $f_0$ of the recording.

Aside from the major two $f_0$ extraction algorithms that have been used in recent years, measurement indices that look at different aspects of the voice elicited FFR responses been reported in various studies. For example, when using the autocorrelation method to extract $f_0$ from the FFR recording, a handful of indices were suggested by Krishnan et al. (2004, 2005). One of them is pitch-tracking accuracy, which is measured
by the time lag associated with the autocorrelation maxim, i.e., the crosscorrelation coefficient between the $f_0$ contour of the stimuli and FFR recording. It represents the degree to which the FFR signal follows the $f_0$ contour of the stimulus. In other words, it gives an estimate of the overall faithfulness of the $f_0$ tracking ability preserved in the FFR recording. Another index was pitch strength, which is measured by average autocorrelation magnitude, takes into account not only the energy at the $f_0$ frequency range but also the energy in the harmonics. It represents the level of the FFR signal in response to the frequency energy in the stimulus. Pitch salience, another FFR measurement index introduced by Bidelman et al. (2009), is also an index that takes the energy in the harmonics into account. It was estimated from each spectra using harmonic templates whereby a series of sieves selected spectral activity at a given frequency and its integer multiples. Each sieve template, representing a signal pitch, was composed of frequency bins located at an $f_0$ and its integer multiples. In essence, pitch salience utilizes the classic pattern recognition model in pitch perception, in which a central processor matches harmonic information contained in the neural response to an internal template in order to compute the heard pitch (Goldstein, 1973; Krishnan et al., 2012). When all the sieve templates are considered, for each template, the degree of pitch salience is estimated by dividing the mean density of activity falling within the sieve bins by the mean density of activity in the whole spectrum, thus providing a contrast between pitch-related activity and the background noise. Jeng, Hu, Dickman, Lin, et al., (2011) and Jeng, Hu, Dickman, Montgomery-Reagan, et al., (2011) presented several other indices, based on both the autocorrelation and the “spectral-density-maxima” $f_0$ extraction
algorithms. For example, frequency error was calculated by finding the absolute
Euclidian distance between the extracted f0 contour from the FFR recording and that of
the stimulus and averaging the errors across the entire recording. It represses the accuracy
of pitch-encoding during the course of stimulus presentation (Jeng, Hu, Dickman, Lin, et
al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011). Slope error indicates
the extent to which the shape of the pitch contours is preserved in the brainstem. To
derive slope error, we first estimate the slope of the regression line of a stimulus f0
contour on an f0-versus-time plot and then subtracting the slope estimate of the stimulus
f0 contour extracted from a recording (Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu,
Dickman, Montgomery-Reagan, et al., 2011). Similar process gives the tracking accuracy
index, which is obtained by conducting linear regression on a recording-versus-stimulus
f0 contour plot and finds the regression r value. It represents the overall faithfulness of
pitch tracking between the stimulus and response f0 contours (Jeng, Hu, Dickman, Lin, et
al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011). The last two indices
used in their studies were the RMS amplitude and the f0 amplitude. The RMS amplitude
is calculated from the root-mean-squared amplitude of the entire recording. It represents
the response amplitude in the time domain. The f0 amplitude represents the amount of
energy located at the response f0 and was calculated by averaging the spectral
amplitudes, obtained by Fourier analysis, along the f0 contour across the temporal axis
(Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al.,
2011).
As discussed above, neural representation of sound in the auditory brainstem is an objective measure of auditory processing and as such can be applied to research and clinical assessment whenever auditory processing is of interest. Voice pitch elicited FFR provides an objective neural metric for determining the effectiveness of remediation strategies, providing the outcome measures that clinicians need to strengthen their role in advocating for auditory training and remediation across the lifespan (Skoe & Kraus, 2010). It has proved its potentials to be an excellent scientific and clinical tool to assess certain aspects of auditory function in the brainstem area. The next major advancement involves the development of algorithms and procedures to make the test of voice pitch elicited FFR automated, independent, objective and reliable. Exploration and development of such automated procedure are particularly important if future researchers and clinicians are to apply the above-mentioned FFR techniques on populations who cannot provide reliable feedback, such as newborn infants or children with communication deficits.
Chapter 3: Methodology

Temporal Procedure

The purpose of this part of study was to explore and evaluate whether a response detection algorithm that is based upon the Pitch Variance Ratio (PVR) index, derived from the Residue Noise Level (RNL) of the temporal property of the recorded waveforms, can detect the presence of an FFR with a satisfactory level of accuracy and efficiency. We believed that the temporal algorithm based on the PVR index can detect the presence of an FFR with a satisfactory level of accuracy and efficiency.

Participants. Thirty college aged adult participants, mean age of 25.4 years, were recruited. All participants, 13 males and 17 females, were recruited in Athens Campus of Ohio University and were native speakers of Mandarin Chinese. Before each test session, the participant was screened by an audiological test to fulfill the requirement of hearing sensitivity ≤ 25 dB HL at octave frequencies from 125 to 8000 Hz. Experimental protocols and procedures used in this study were approved by the Ohio University Institutional Review Board.

Stimulus preparation. Four monosyllabic Mandarin Chinese syllables with different pitch contours (Tone 1 flat /yi1/ 'clothes', Tone 2 rising /yi2/ 'aunt', Tone 3 dipping /yi3/ 'chair', Tone 4 falling /yi4/ 'meaning') were prepared. These stimulus tokens were recorded in a sound booth from a male speaker whose native language was Mandarin Chinese. An Audio-technica AT825 microphone, connected through a preamplifier and an analog-to-digital converter, was placed in the booth to pick up the sound. The recorded voice tokens were sampled digitally using the BLISS v7 (Brown
Lab Interactive Speech System, Providence, RI) at a 40 kHz sampling rate. These voice samples then underwent a temporal normalization to a unified duration of 250 ms with a rising/falling time of 10ms. The final stimulus tokens of the four Mandarin syllables had the frequency ranges containing their $f_0$ contours as follows: Tones 1: 163-180 Hz, Tone 2: 116-157 Hz, Tone 3: 98-125 Hz and Tone 4: 105-156 Hz.

**Stimulus presentation.** A LabVIEW (National Instruments, Austin, TA) script was used to control the presentation of the stimulus tokens and trigger synchronization. The stimulus tokens were presented through a 16-bit digital-to-analog converter (National Instruments, PCI 6221), a Low-pass filter (WaveTek Model 442, low-pass filter with a cut-off frequency at 20k Hz), an attenuator (Tucker Davis Technologies model PA4) and a headphone buffer (Tucker Davis Technologies model HB6). The acoustic tokens were presented monaurally to the participant's right ear through a custom-built electromagnetically-shielded insert earphone modified from an ER3-A earphone (Bio-logic Systems Corp., Mundelein, Illinois). The presented sounds were calibrated using a Larson Davis System 824 sound level meter (Larson Davis system, Depew, New York). Each stimulus token were presented at 70 dB SPL for blocks of 2000 accepted sweeps for each token. The inter-stimulus interval between adjacent stimulus tokens was set at 45 ms. The four Mandarin tones were presented in a random order within and across participants. To ensure that the stimulus artifact was eliminated from the recordings appropriately, in each testing session, a control condition (sound tube occluded and removed from the participant’s ear) was conducted at the end of each recording session to provide EEG wave forms that have no direct relationship to the stimulus presentation.
**Recording Parameters.** Three gold-plated recording electrodes (GRASS System of Viasys Healthcare, Madison, WI) were applied to each participant on the forehead (Fpz, non-inverting), right mastoid (M2, inverting), and left mastoid (M1, ground). All electrode impedances were maintained under 3 kOhms at 10 Hz. Continuous recordings were amplified using an Opti-Amp amplifier with a gain of 50k, online-filtered 100-3000 Hz 6 dB/octave (Intelligent Hearing Systems, Miami, FL), and digitized at a rate of 20k samples per second. The continuous recordings were stored on a computer for offline analysis.

**Data analysis and interpretation.** All data was arranged and analyzed using MATLAB (MathWorks, Natick, MA) and EEGLab 6.01b (Swartz Center for Computational Neuroscience, San Diego, CA). In order to isolate spectral energies at the frequency range near the \( f_0 \) contours, continuous data was digitally filtered using a brick-wall, linear-phase finite-impulse-response (FIR) band-pass filter (85-1500 Hz, 500th order). Filtered recordings were segmented into sweeps of 295 ms in length. An artifact rejection limit of \( \pm 25 \mu V \) was used to ensure the quality of the recorded waveforms. For each condition, a total of 2000 accepted recording sweeps were collected.

This part of this study utilized an algorithm that was based upon the residue noise level (RNL) index of the temporal property of the recorded waveform (Schimmel, 1967; Wong & Bickford., 1980). Briefly, this algorithm was implemented by first dividing all recorded sweeps into two separate banks - with all the odd-number sweeps stored in one bank and even-number sweeps in the other. Assuming each sweep contains a representation of the desired neural responses, the average of the two banks of sweeps
would then represent the response waveform itself, while the difference would represent
the noise. The average of the two banks essentially is the same as the average of all
sweeps collected, and thus can be expressed using Formula 3.

\[
\overline{A_N(t)} = \frac{1}{N} \left[ \sum_i \{EEG_i(t)\} \right], \quad (3)
\]

where \(A_N\) represents the averaged signal between the two banks, \(N\) is the number of
sweeps, \(EEG\) indicates the recorded waveform for each sweep, and \(i\) denotes each
individual sweep ranging from 1 to \(N\).

Similarly, the difference between the two banks can be expressed using Formula
4.

\[
\overline{D_N(t)} = \frac{1}{N} \left[ \sum_i \{EEG_i(t) \cdot (-1)^i\} \right] \quad (4)
\]

where \(D_N\) represents the difference between the two banks, \(N\) is the number of sweeps,
\(EEG\) indicates the recorded waveform for each sweep, and \(i\) denotes each individual
sweep ranging from 1 to \(N\).

Alternatively, each recording can be constructed by two components:

\[
\overline{A_N(t)} = s(t) + \overline{n(t)} \quad (5)
\]

where \(s(t)\) represents the physiological response and \(n(t)\) represents noise. With
increasing number of collected sweeps, the averaged response \(\overline{A_N(t)}\) would approach the
real waveform of the elicited response \(s(t)\). On the other hand, the difference between the
two banks (i.e., the noise \(\overline{D_N(t)}\)) would optimally approximate the true value of the noise
\(n(t)\). Thus, if the recorded waveform were to contain an elicited neural response, which
should then be temporal locked to the stimulus in each recorded sweep, the response
component \(s(t)\) should not contribute in \(\overline{D_N(t)}\), because it represents response and
should have been cancelled out during the alternation of addition and subtraction. Thus, we would have two estimates that would approximate the elicited neural response $A_N(t)$ and noise $D_N(t)$, which is of the order of $1/\sqrt{N}$. 
Figure 11. Flow chart of the Pitch Variance Ratio based automated procedure for detecting the presence or absence of the frequency-following response to voice pitch. This automated procedure includes several procedural steps: (1) present stimulus tone and record continuous EEGs, (2) separate the recorded EEGs into two banks that stored odd-number and even-number sweeps, respectively, (3) perform addition and subtraction of the averaged waveforms stored in the two banks, (4) calculate Pitch Variance Ratio (PVR) value, (5) compare PVR value with the response detection criterion (i.e., the $PVR_{\text{critical}}$ value), and (6) accept or reject the presence of an FFR.
To determine the presence/absence of a response signal, computation of an appropriate signal-to-noise ratio (SNR) is almost inevitable (see Figure 11). Because variance is in essence a measure of the overall amplitude of a response signal, the ratio of the variance of the response signal (i.e., $A_N(t)$) to the variance of the noise (i.e., $D_N(t)$) was used to indicate the magnitude that the response signal stood out from the noise (i.e., SNR). This concept can be expressed using Formula 6.

$$SNR = \frac{\text{Var}(A_N(t))}{\text{Var}(D_N(t))} \tag{6}$$

where SNR stands for signal-to-noise ratio, $\text{Var}(A_N(t))$ represents the variance of the average of the two banks, $\text{Var}(D_N(t))$ indicates the variance of the difference between the two banks. If we continue to break down the SNR calculation, we have

$$SNR = \frac{\text{Var}(A_N(t))}{\text{Var}(D_N(t))} = \frac{\text{mean power of } s(t)}{\text{mean power of } n(t)} + \frac{\text{mean power of } n(t) \text{ in } A_N(t)}{\text{mean power of } n(t) \text{ in } D_N(t)} \tag{7}$$

Denote the second term in Formula 7 by $F$ and using $\sigma^2(n)/N$ as the denominator of the first term, we have

$$SNR \approx N \cdot \frac{\sigma^2(s)}{\sigma^2(n)} + F = Nf^2 + F \tag{8}$$

where $f$ is the amplitude ratio of $s(t)$ and $n(t)$, i.e., the real SNR. Thus, when there is no response, where $s(t)=0$, and $f=0$, we have $SNR=F$, which is the $F$ ratio of the noise. When there is a response, then $SNR \approx Nf^2 + 1$. The resulting SNR level is called the Residue Noise Level (Wong & Bickford, 1980).
To implement the RNL algorithm on the automated response detection procedure for the FFR to voice pitch, the averaged recording \( \bar{A}_N(t) \) first went through a cross-correlation with the stimulus waveform to identify the time shift that produces the maximum cross-correlation value between the 3-10 ms response window. A 250 ms segment of data was then extracted from the averaged recording (starting from the maximum cross-correlation value) and was denoted as \( \bar{A}_N(t)_c \). The same procedure was applied to the estimate of the noise and the result was denoted as \( \bar{D}_N(t)_c \). Thus, for the realm of the FFR to voice pitch, the magnitude of the voice-pitch elicited response can be expressed in the form of SNR, which takes into account of the variance ratio between the estimated response and noise. In this study, the term pitch variance ratio (PVR) is introduced to indicate the SNR of each recording and will be used to determine the presence or absence of an FFR. In sum, the PVR can be expressed as Formula 10.

\[
PVR = \frac{\text{Var}(\bar{A}_N(t)_c)}{\text{Var}(\bar{D}_N(t)_c)}
\]  

(10)

For each recording, such a calculation was performed and a PVR value was obtained. Because the process of calculating a PVR index is in essence an \( F \)-test of equal variance between the estimated signal and noise, a \( PVR_{\text{critical}} \) value can be determined by the degrees of freedom of the two variances and the desired \( \alpha \) level to set a criterion for the automated procedure. The underlying logic of the automated procedure was to use the \( PVR_{\text{critical}} \) value as a response detection criterion. Any recording that has a PVR value greater than the \( PVR_{\text{critical}} \) value will be considered to contain physiological energy that stands out from the background noise, thus will be considered to contain a voice pitch elicited FFR. Apparently the \( PVR_{\text{critical}} \) value will vary depending on the choice of \( \alpha \) level.
and the degrees of freedom of the estimated signal and noise. For our FFR recordings, the degree of freedom is a constant (i.e., 5000 - 1 = 4999) for both the estimated signal and noise. If α level is set at 0.05, an $F$-statistic table gives a PVR\text{critical} value of 1.05. On the other hand, if α level is set at 0.10, an $F$-statistic table gives a PVR\text{critical} value of 1.04. In this study, the more conservative PVR\text{critical} value of 1.05 was used as the detection criterion in the automated response detection procedure. To better illustrate the concepts mentioned above, the entire process of the automated response detection procedure is demonstrated in Figure 11.

**Statistical analysis.** One-way repeated measure ANOVA was conducted between the experimental conditions and the control condition to test the null hypothesis that PVR values obtained from the experimental conditions are not significantly different than those obtained from the control condition.

**Effect of temporal progression and number of sweeps.** Voice pitch elicited FFR has a unique feature that the response would follow the temporal progression of the $f_0$ of the stimulus so that the temporal progression property of FFR is as important as the overall response property. To use PVR to examine such temporal progression property, each recording was divided into ten temporal segments: From the very first data point of each recording length, every 500 data points were marked as one temporal block. Since each recording waveform is 250 ms in length, with a sample rate of 20 kHz/s, the total data points is then 5000 and that resulted in a total of ten temporal blocks within each recording waveform. For each temporal block, the above described PVR calculation
algorithm was carried out to examine the temporal progression of PVR within the recording.

To further explore the efficiency of the automated procedure, PVR values for all the experimental and control conditions were calculated as a function of number of sweeps. By gradually increasing the number of sweeps included in the averaged waveform (starting from the first recorded sweep), all recordings were reconstructed and re-calculated using the same data analysis procedures as mentioned above (including the averaging, filtering, segmentation, cross-correlation with the stimulus token, and the PVR computation).

**Spectral Procedure**

The purpose of this part of the study was to explore and evaluate whether response detection algorithm that is based upon the statistic properties of the spectral energy distribution in the recorded waveforms can detect the presence of an FFR with a satisfactory level of accuracy and efficiency. We believed that the statistics-amended spectral response extraction algorithm can detect the presence of an FFR with a satisfactory level of accuracy and efficiency.

**Data analysis and interpretation.** For each recording session, same as the procedure used in the first part of this study, a total of 2000 accepted sweeps were collected. The averaged recording then underwent a cross-correlation process with the stimulus token to identify the time shift that produces the maximum cross-correlation value between the 3-10 ms response window. A 250 ms segment of data was then drawn out from the original averaged waveform, starting from the calculated maximum cross-
correlation value. With a sampling rate of 20k samples/sec and a length of 250 ms, this data segment consists 5000 data points. A time-varying spectral representation, i.e., a spectrogram, was calculated to present the spectral density distribution and its temporal variation.
Table 1

*Fundamental Frequencies (f0) of the Mandarin Chinese Syllables*

<table>
<thead>
<tr>
<th>Tone</th>
<th>Time (ms)</th>
<th>1</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
<th>190</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>163</td>
<td>164</td>
<td>165</td>
<td>167</td>
<td>172</td>
<td>175</td>
<td>177</td>
<td>179</td>
<td>180</td>
<td>179</td>
<td>179</td>
<td>177</td>
<td>175</td>
<td>174</td>
<td>174</td>
<td>175</td>
<td>179</td>
<td>180</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>117</td>
<td>118</td>
<td>118</td>
<td>117</td>
<td>116</td>
<td>117</td>
<td>119</td>
<td>124</td>
<td>127</td>
<td>129</td>
<td>136</td>
<td>142</td>
<td>144</td>
<td>146</td>
<td>152</td>
<td>155</td>
<td>156</td>
<td>157</td>
<td>157</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>113</td>
<td>113</td>
<td>111</td>
<td>105</td>
<td>103</td>
<td>101</td>
<td>99</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>105</td>
<td>109</td>
<td>113</td>
<td>116</td>
<td>119</td>
<td>123</td>
<td>124</td>
<td>124</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>186</td>
<td>185</td>
<td>185</td>
<td>184</td>
<td>183</td>
<td>179</td>
<td>176</td>
<td>174</td>
<td>171</td>
<td>169</td>
<td>167</td>
<td>165</td>
<td>162</td>
<td>160</td>
<td>156</td>
<td>146</td>
<td>144</td>
<td>142</td>
<td>141</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Extended from the “narrow-band spectrogram” algorithm (Jeng, Hu, Dickman, Lin, et al., 2011), a distinctive statistics-amended response detection algorithm was used on the averaged EEG waveform and its corresponding spectral density distribution representation. Unlike the “narrow-band spectrogram” algorithm, which uses solely the information from the spectral-domain and searches for the maximal spectral density in a predetermined frequency range, the new statistics-amended spectral algorithm utilize both information from the spectral energy density distribution and the statistic property of such distribution. First off, the way that the new statistics-amended spectral algorithm extracts the \( f_0 \) contours differs from the old one: the algorithm uses the information embedded in the data to determine the stimulus that was used in the recording session (Jeng, Hu, Dickman, Lin, et al., 2011). With the stimulus (one of the four Mandarin syllables) information, the algorithm first read the pre-stored \( f_0 \) contour array of the stimulus (see Table 1). Similar to the “narrow-band spectrogram” algorithm, the new algorithm starts searching for the maximal spectral density value in each windowed segment. However, instead of doing so in a fixed pre-defined frequency range, it will extend its searching range based upon the frequency value of the stimulus’ \( f_0 \) in that segment. In other words, the searching ranges changes dynamically with the temporal advancement of the recorded waveform. Specifically, within in each segment, the algorithm has a range, \( \pm 5 \) Hz in this study, centered at the frequency value of the stimulus’ \( f_0 \), within which it calculates the mean spectral energy value and it represents the energy of the signal. Outside the signal range, the algorithm also have two ranges that are used to calculate the mean spectral energy for the background noise, on above the
upper frequency limit of signal range, 20 Hz in this study, and one below the lower
frequency limit, 10 Hz in this study. With a spectral resolution of 1 Hz, the pooled noise
part leaves the algorithm a total number of 30 frequency points for the estimation of the
noise. A one-sample Student’s $t$-test is then conducted upon the mean spectral energy
values of the signal and the noise, i.e., the $t$-test is carried out between the signal part and
the combined noise part in every windowed segments. Thus, for each $t$-test, the algorithm
tests the null hypothesis that the averaged energy in noise part is similar to the signal
energy, with a confidence level of 95% and degree of freedom at 29. If a significant
difference is detected by the $t$-test, that windowed segment at the moment will be marked
as a “response” segment; otherwise, it will be marked as a “non-response” segment. The
algorithm then continues along the temporal axis and repeats the above procedures until
the end of the recorded waveform. As there are a total of 201 windowed segments, same
$t$-tests are carried out in all of these segments. The results of these 201 $t$-tests form a
vector of response marks: some would have a mark of “response” and others would be
“non-response”. A response detection index, Relative Significance Level (RSL), was
calculated at this point to be further used in the response detection procedure. For the
windowed segments that were marked as a “response” segment, they were given a value
of 1, whereas those marked “non-response” were given a value of 0. The RSL value was
calculated as

$$RSL = \sum_{i=1}^{201} Segment_i$$

(11)

where the value of Segment is either 1 or 0, depending on the result of the $t$-test in that
windowed segment. Thus, a strong voice pitch elicited FFR recording would have a
perfect RSL score of 201, while the RSL value obtained from a control condition, i.e., no response, would have a score that was close to 0.

**Figure 12.** Flow chart of the Relative Significance Level based automated response detection procedure. It involves several procedural steps including (1) reading, (2) addition and subtraction of the averaged waveforms of the two banks, (3) cross-correlation with the stimulus waveforms, (4) calculation of Pitch Variance Ratio (PVR), and (5) comparison with the response detection criterion (i.e., $PVR_{critical}$ value).

The response detection criterion that was used in the automated procedure was determined by selecting a desired RSL score ($RSL_{critical}$), i.e., the percentage of “significance” segments in the 201 total segments. The automated response detection procedure uses the criterion as a decision maker: if a recording provides a RSL value that
is greater than $RSL_{\text{critical}}$, it would be considered to contain a voice pitch elicited FFR, otherwise it would be considered as a recording that contains no physiological response (see Figure 12).

**Statistical analysis.** Paired $t$-tests were carried out between designated signal sections and designated noise sections in the spectral density distribution representation, to obtain the RSL values. One-way repeated measure ANOVA was conducted between the experimental conditions and the control condition to test the null hypothesis that RSL values obtained from the experimental conditions are not significantly different than those obtained from the control condition.

**Effect of temporal progression and number of sweeps.** To further explore the efficiency of the statistics-amended spectral response detection algorithm, we examined the distribution of the RSL values obtained from the four tones and the control condition by increasing the number of sweeps that were included in the averaged waveform: Each recording waveform was examined from an evolving perspective: Every 200 sweeps, one averaging process was carried out over the entire recording time waveform and the response percentage calculation was performed on the resulting averaged waveform.

**Comparisons with Human Judgments**

The purpose of this part of the study was to explore whether the automated procedures that are developed from the PVR-based temporal algorithm and statistics-amended spectral algorithm could derive comparable ROC curves, when compared with human judgments. We believed that the automated procedures that are developed from
the RNL-based temporal algorithm and statistics-amended spectral algorithm could derive comparable ROC curves as compared to those derived from human judgment.

**Data analysis and interpretation.** The output from the above mentioned response detection algorithms were compared with subjective human judgment. Since the applications of FFR have yet to be used widely in research and clinic settings, there are no normative data base or response templates that we could rely on. Further, as of the properties for almost every electrophysiological measurement, the subjective interpretations from an experienced operator are still considered the “gold standard.” Thus, we propose to have three experienced persons to read and interpret the recorded waveforms and make the final judgments as whether the waveform could be considered as a positive voice pitch elicited FFR or a recording contains no biological response.

Three experienced judges were asked to complete the task. To ensure the level of expertise of the judges, they have received substantial amount of trainings of conducting electrophysiological measurements, interpreting collected data and reading FFR spectrograms. A MATLab script was used to facilitate the interpretation process: all the collected FFR recordings, including the recordings using all four Mandarin syllables and the control recordings, were presented at a random order to the judge. Each judge was observing each spectrogram of the recording while making reference to the spectrogram of the stimulus that was used in that specific recording session. A two-alternative-forced-choice paradigm was used when the judge was determining the quality of each recording: there were only two buttons (“Yes” and ”No”) on the screen for the judge to choose from. Same protocol was applied to all three judges. For each recorded waveform, if two out of
three judges had the same decision (i.e., presence or absence of a voice pitch elicited FFR), that decision was recorded as the final judgment and was stored to be further compared to the judgment made by the automated procedure.

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Human Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
</tr>
<tr>
<td><strong>Positive</strong></td>
<td>True Positive (TP)</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>False Negative (FN)</td>
</tr>
</tbody>
</table>

**Operating characteristics:**

Sensitivity (Power) = \( \frac{TP}{TP+FN} \) %

Specificity = \( \frac{TN}{FP+TN} \) %

False-Negative Rate = \( \frac{FN}{TP+FN} \) %

False-Positive Rate = \( \frac{FP}{FP+TN} \) %

Positive-Predicative Value = \( \frac{TP}{TP+FP} \) %

Negative-Predictive Value = \( \frac{TN}{FN+TN} \) %

Efficiency = \( \frac{TP+TN}{TP+FP+TN+FN} \) %

Type I Error (Alpha) = False-Positive Rate = (100 - specificity) %

Type II Error (Beta) = False-Negative Rate = (100 - sensitivity) %

*Figure 13.* A matrix describing the calculation of the operating characteristics.
For both the PVR based temporal procedure and the RSL based spectral procedure, the output of the procedures were compared with decisions made by human judges to construct the Receiver Operating Characteristics (ROC). For the PVR based temporal procedure, as the response detection criterion was pre-determined by statistics ($\alpha = 0.05$, $PVR_{critical} = 1.05$), the comparisons between the procedure and human judgments were carried out once to reveal the sensitivity and specificity for the four tones. For the RSL based spectral procedure, the response detection criterion was based on the desired $RSL_{critical}$ value. Thus, a ROC space was constructed based upon the comparisons between the procedure and human judgments: the desired percentage of significant segments, i.e., the $RSL_{critical}$ values, was gradually increased from 5% to 100%, with a step size of 5%. For each desired percentage value ($RSL_{critical}$), a comparison between the output of the procedure and the human judgment was carried out to find elements that are need to construct the ROC space. For recordings that receive positive decision from both the procedure and the human judge, they will be marked as True Positives. The rest three elements: False Positives, False Negatives and True Negatives, will be determined in the same manner. These elements were recorded for each $RSL_{critical}$ value, for all four tones. Further evaluation and discussion of the effectiveness, the efficiency and other properties, such as the sensitivity and specificity, of the procedure was conducted using this ROC space (see Figure 13).
Figure 14. Pitch Variance Ratio (PVR) values obtained in the experimental and control conditions. Group comparison (panel A) shows that the PVR values obtained in the four experimental conditions were significantly larger than those obtained in the control condition. Comparison for each individual tone (panel B) shows the PVR values obtained from each individual tone was also significantly larger than that of the control condition. Red horizontal lines indicate the mean PVR value for each condition, blue lines indicate the medians. The upper and lower boundaries of the box indicate the 25th and 75th percentiles, respectively. Whisker bars indicate the 10th and 90th percentile.
Table 2

*Temporal Progression of the Pitch Variance Ratio Values (Mean)*

<table>
<thead>
<tr>
<th>Tone</th>
<th>Time bins</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2.57</td>
</tr>
<tr>
<td>2</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>1.91</td>
</tr>
<tr>
<td>4</td>
<td>2.17</td>
</tr>
<tr>
<td>Ctrl</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Chapter 4: Results

Temporal Procedure

**Waveforms of voice pitch elicited FFR.** Figure 13 demonstrates typical examples of the time waveforms obtained in one participant (Participant 25). The addition of the two banks (odd-number and even-number of sweeps, Figure 2, top left and bottom left) provides waveforms provides the averaged signal (Figure 13, top right), while the subtraction renders an estimation of the noise (Figure 13, bottom right).

**Distinctive distributions of the PVR values in voice-pitch elicited FFRs.** The PVR values (see Table 2) obtained in both the experimental and control conditions are shown in Figure 14. A group comparison (Figure 14A) between the experimental (Tones 1, 2, 3, and 4 combined) and control conditions revealed that the PVR values obtained from the experimental conditions were significantly larger than those obtained in the control condition \( (p < 0.001) \). The PVR values obtained from the four experimental conditions had a mean of 1.738 (SD = 1.651), whereas the PVR values obtained from the control condition had a mean of 0.261 (SD = 0.042). Note the mean PVR value was approximately 6.7 times larger than that of the control condition. For the control condition, its PVR values were clustered within a much smaller range than that of the experimental conditions \( (p < 0.001) \).

For each individual tone (see Figure 14B), the mean of its corresponding PVR values (Tones 1, 2, 3 and 4 mean \( \pm SD = 1.799 \pm 1.597, 1.134 \pm 0.858, 2.550 \pm 2.221 \) and 1.468 \( \pm 1.351 \), respectively) was also significantly larger than that of the control condition \( (p < 0.001 \text{ for Tones 1, 3 and 4, } p = 0.031 \text{ for Tone 2}) \). A *post hoc* Tukey-
Cramer analysis showed no statistical difference among the four tones with the exceptions that Tones 3 versus Tone 2 \((p < 0.001)\) and Tones 3 versus Tone 4 are significantly different from each other \((p = 0.004)\).

*Figure 15.* Temporal progression of the Pitch Variance Ratio values along the temporal axis for all four stimulated conditions and the control condition. Each temporal block represents 25 ms in the recorded waveform.
Table 3

Pitch Variance Ratio Values as a Function of the Number of Sweeps

<table>
<thead>
<tr>
<th>Tone</th>
<th>Number of Sweeps (200)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.42</td>
<td>0.55</td>
<td>0.70</td>
<td>0.85</td>
<td>0.99</td>
<td>1.13</td>
<td>1.27</td>
<td>1.44</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.34</td>
<td>0.42</td>
<td>0.48</td>
<td>0.57</td>
<td>0.65</td>
<td>0.75</td>
<td>0.86</td>
<td>0.95</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.49</td>
<td>0.78</td>
<td>1.03</td>
<td>1.18</td>
<td>1.38</td>
<td>1.62</td>
<td>1.80</td>
<td>2.05</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.40</td>
<td>0.52</td>
<td>0.66</td>
<td>0.74</td>
<td>0.85</td>
<td>0.99</td>
<td>1.12</td>
<td>1.23</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>Ctrl</td>
<td>0.24</td>
<td>0.25</td>
<td>0.24</td>
<td>0.23</td>
<td>0.22</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Temporal progression of PVR values. To examine the temporal progression of the PVR values for all the experimental and control conditions, the mean PVR values for each time bin were collected and organized progressively in the temporal domain (see Table 2). In each time bin (25 ms in length, with 500 data points), the same data construction and analysis procedures were used as mentioned above (including the averaging, filtering, segmentation, cross-correlation with the stimulus token, and the PVR computation).

Figure 15 shows the temporal progression of the PVR values. Within each tone condition, the PVR values obtained from all the Tones 1, 2, 3 and 4 remained relatively stable along the temporal progressions. (No significant different were found between the time bins in each tone). When comparing to the control condition, all tones had significantly larger mean PVR values than those obtained in control conditions for every time bin along the temporal progression (p < 0.05) except for the first one. When compared between the four tones, Tone 3 had several time bins that produce significantly larger mean PVR values (p < 0.05) than other three tones (time bin number 3, 4, 5, 7 and 8). The rest of tones, Tone 1, 2 and 4, had similar PVR values in their corresponding time bins, when compared to each other.

Effect of number of sweeps on PVR. To further explore the efficiency of the automated procedure, PVR values for all the experimental and control conditions were calculated as a function of number of sweeps (see Table 3). By gradually increasing the number of sweeps included in the averaged waveform (starting from the first recorded sweep), all recordings were reconstructed and re-calculated using the same data analysis
procedures as mentioned above (including the averaging, filtering, segmentation, cross-correlation with the stimulus token, and the PVR computation).

Figure 16. Pitch Variance Ratio values as function of the number of sweeps. Note the PVR values obtained in the experimental conditions (Tones 1, 2, 3 and 4) all increase with increasing number of sweeps, while the PVR values obtained in the control condition remained relatively small and stable over the time course of data collection.

Figure 16 shows the PVR trends as a function of number of sweeps. The PVR values obtained from all the Tones 1, 2, 3 and 4 increased with increasing number of sweeps, whereas the PVR values obtained in the control condition remained relatively stable (0.241 at 200 sweeps and 0.261 at 2000 sweeps).
In the automated response detection procedure, the PVR value will constantly be updated with increasing number of sweeps until it exceeds the response detection criterion \( \text{PVR}_{\text{critical}} = 1.05 \). Tones 1, 2, 3 and 4 exceeded the \( \text{PVR}_{\text{critical}} \) criterion at 1200, 2000, 600 and 1400 sweeps, respectively. Among the four tones used in this study, Tone 3 required the least number of sweeps to reach the response detection criterion.
### Table 4

**Relative Significance Levels as a Function of the Number of Sweeps**

<table>
<thead>
<tr>
<th>Tone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.68</td>
<td>0.73</td>
<td>0.81</td>
<td>0.86</td>
<td>0.88</td>
<td>0.89</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>0.60</td>
<td>0.70</td>
<td>0.77</td>
<td>0.81</td>
<td>0.86</td>
<td>0.88</td>
<td>0.91</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>0.73</td>
<td>0.78</td>
<td>0.82</td>
<td>0.83</td>
<td>0.85</td>
<td>0.85</td>
<td>0.86</td>
<td>0.87</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>4</td>
<td>0.69</td>
<td>0.76</td>
<td>0.81</td>
<td>0.82</td>
<td>0.83</td>
<td>0.85</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>C</td>
<td>0.43</td>
<td>0.41</td>
<td>0.37</td>
<td>0.36</td>
<td>0.34</td>
<td>0.35</td>
<td>0.33</td>
<td>0.31</td>
<td>0.27</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Figure 17. Relative Significance Level values obtained in the experimental and control conditions. Group comparison (panel A) shows that the RSL values obtained in the four experimental conditions were significantly larger than those obtained in the control condition. Comparison for each individual tone (panel B) shows the RSL values obtained from each individual tone was also significantly larger than that of the control condition. Red horizontal lines indicate the mean RSL value for each condition, blue lines indicate the medians. The upper and lower boundaries of the box indicate the 25th and 75th percentiles, respectively. Whisker bars indicate the 10th and 90th percentile.

Spectral Procedure

RSL distribution. The RSL values obtained in both the experimental and control conditions are shown in Figure 17. A group comparison (see Figure 17A) between the experimental (Tones 1, 2, 3, and 4 combined) and control conditions revealed that the RSL values obtained from the experimental conditions were significantly larger than those obtained in the control condition ($p < 0.001$). The RSL values obtained from the
four experimental conditions had a mean of 0.91 (SD = 0.11), whereas the RSL values obtained from the control condition had a mean of 0.24 (SD = 0.25). Note the mean RSL value was approximately 4 times larger than that of the control condition. For the control condition, its RSL values were clustered within a much smaller range than that of the experimental conditions ($p < 0.001$).

For each individual tone (see Figure 17B), the mean of its corresponding RSL values (Tones 1, 2, 3 and 4 mean ± SD = 0.93 ± 0.11, 0.93 ± 0.09, 0.90 ± 0.09 and 0.88 ± 0.13, respectively) was also significantly larger than that of the control condition ($p < 0.001$ for all four tones). A post hoc Tukey-Cramer analysis showed no statistical difference among the four tones.

![Figure 18](image)

*Figure 18.* Relative Significance Level values as function of the number of sweeps. Note the RSL values obtained in the experimental conditions (Tones 1, 2, 3 and 4) all increase with increasing number of sweeps, while the RSL values obtained in the control condition remained relatively small and decreased over the time course of data collection.
**Effect of number of sweeps on RSL.** To further explore the efficiency of the automated procedure, RSL values for all the experimental and control conditions were calculated as a function of number of sweeps (see Table 4). By gradually increasing the number of sweeps included in the averaged waveform (starting from the first recorded sweep), all recordings were reconstructed and re-calculated using the same data analysis procedures as mentioned above (including the averaging, filtering, segmentation, cross-correlation with the stimulus token, and the RSL computation).

Figure 18 shows the RSL trends as a function of number of sweeps. The RSL values obtained from all the Tones 1, 2, 3 and 4 increased with increasing number of sweeps, whereas the RSL values obtained in the control condition remained relatively small and decreased significantly ($p < 0.05$) over increasing number of sweeps (0.43 at 200 sweeps and 0.24 at 2000 sweeps). When comparing to the control condition, all tones had significantly larger mean RSL values than those obtained in control conditions ($p < 0.05$), starting from the very first 200 sweeps and the difference became progressively larger as the number of sweeps increases. When compared between the four tones, the RSL values obtained in each sweep increment were close to each other, with no significant difference between the four tones.

**Comparisons With Human Judgments**

To evaluate the effectiveness of the two automated response detection procedures, the output of the two procedures were compared with judgments that were made by experienced researchers.
Figure 19. Receiver Operating Characteristics of the Pitch Variance Ratio based temporal procedure. Marked points represent the comparison between the outputs from Pitch Variance Ratio based procedure and human judgments. The abscissa represents the False Positive Rate (FPR), i.e., 1-specificity, whereas the ordinate represents the True Positive Rate (TPR), i.e., sensitivity.

Figure 19 demonstrates the Receiver Operator Characteristics (ROC) space with results obtained from the outputs from PVR based procedure ($PVR_{critical} = 1.05$) compared with human judgments. Specifically, results from PVR based procedure provides sensitivity (true positive rate) of 42.31%, 62.96%, 86.21% and 54.55%, for Tone 1, 2, 3 and 4, respectively. For specificity, it was recorded that the PVR based procedure provided numbers of 75%, 66.67%, 100% and 50% for Tone 1, 2, 3 and 4, respectively.
Table 5

*True Positive Rate of the Relative Significance Level Based Spectral Procedure*

<p>| Acceptance (%) | 5  | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|----------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| Tone           | 1  | 100| 100| 100| 100| 100| 100| 100| 100| 100| 100| 100| 100| 93 | 61 | 36 |
| 2              | 100| 100| 100| 100| 100| 100| 100| 100| 100| 100| 97 | 97 | 93 | 90 | 77 | 60 | 40 |
| 3              | 100| 100| 100| 100| 100| 100| 100| 100| 100| 100| 97 | 97 | 90 | 79 | 69 | 45 | 7  |
| 4              | 100| 100| 100| 100| 100| 100| 100| 100| 100| 96 | 96 | 96 | 89 | 85 | 85 | 78 | 70 | 33 | 7  |</p>
<table>
<thead>
<tr>
<th>Acceptance (%)</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
<th>85</th>
<th>90</th>
<th>95</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>1/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>91</td>
<td>84</td>
<td>72</td>
<td>66</td>
<td>50</td>
<td>38</td>
<td>22</td>
<td>19</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tone 2</td>
<td>2/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>83</td>
<td>70</td>
<td>63</td>
<td>47</td>
<td>33</td>
<td>20</td>
<td>17</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tone 3</td>
<td>3/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90</td>
<td>84</td>
<td>71</td>
<td>65</td>
<td>48</td>
<td>35</td>
<td>23</td>
<td>19</td>
<td>13</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tone 4</td>
<td>4/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>91</td>
<td>85</td>
<td>73</td>
<td>67</td>
<td>52</td>
<td>39</td>
<td>27</td>
<td>24</td>
<td>15</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 20. Receiver Operating Characteristics of the Relative Significance Level based spectral procedure. Marked points represent the comparison between the outputs from Relative Significance Level based procedure and the human judgments. The abscissa represents the False Positive Rate (FPR), i.e., 1-specificity, whereas the ordinate represents the True Positive Rate (TPR), i.e., sensitivity.

When examining the effectiveness of the RSL based spectral procedure, the True Positive Rate (TPR, i.e., sensitivity) and the False Positive Rate (FPR, i.e., 1-specificity) were calculated for each tone condition for every percentage of significant elements (i.e., \( RSL_{\text{critical}} \) value). The percentages of significant elements were increased from 5% to 100% (see Tables 5 and 6). Based on these TPR and FPR values, Receiver Operator Characteristic curves were constructed for the four tones (see Figure 20). It seemed that at around 65-75% acceptance rate, the RSL based spectral procedure reached a level that it could provide satisfactory TPR and FPR values, for all four tones.
Chapter 5: Discussion

This study developed and evaluated automated response detection procedures for voice pitch elicited FFR that were based on the statistical properties of the recorded data itself and was independent from subjective human judgments. The PVR based automated procedure used the temporal information of the time waveforms and calculated the statistical property of the data in the time domain, eliminating the need for subjective interpretation or manual input. The RSL based automated procedure took advantage of the unique feature of FFR that a response mimics the spectral properties of the stimulus closely. When calculating the RSL values, statistical tests were performed on these spectral properties and the results were used in the response detection procedure. Previously in the field of voice pitch elicited FFR, most of the judgments of the existence of such a response were made by human observers, which limits the potential future scientific and clinical applications of the FFR. Results obtained from the present study indicated that the automated procedure could be a useful tool for the FFR research.

Current Study

Subjective interpretation and current study attempt to automate the response detection for FFR. Traditionally, the examination and interpretation of FFR recording were made from subjective decisions made by human experimenters. Subjective decision methods are based on one or more individual’s visual judgment that a FFR is present in the averaged signal. In FFR, the experimenter inspects the averaged data, such as waveforms and spectrograms, and visually determines whether there is a distinctive waveform that follows the morphology of the stimulus or discernible energy at
the desired frequency range (usually at or around the $f_0$) (Jeng et al., 2010; Jeng, Hu, Dickman, Lin, et al., 2011; Jeng, Hu, Dickman, Montgomery-Reagan, et al., 2011; Krishnan et al., 2004, 2005). In most cases, these inspections were made by experimenters who have received a substantial amount of trainings and years of experience. If the voice pitch elicited FFR were to see application in clinical settings in the near future, likely as an auditory screening tool to pick out patients at potential risks of developing pitch resolving deficits, subjective and training-intensive response detection process could be supplemented or largely replaced by improved objective response detection procedures.

To address the shortcomings of the visual methods for detecting a voice pitch elicited FFR response and to make the tests truly objective, objective response detection procedures that use statistical techniques are needed. Compared to conventional FFR detection process (Jeng et al., 2010), the automated procedures used in this study eliminate the need for visual inspection of the time waveform and/or the spectral energy, thus would avoid potential biases inherent in subjective decision making processes and reduce the amount of training for the operator.

One step further from the automated algorithm reported by Jeng et al. (2011b), this study made an improvement that both response detection procedures were independent from any subjective human interpretation because this procedure was based the statistical properties of the recorded data and nothing else. Specifically, the PVR based temporal procedure looked at the nature of the voice pitch elicited FFRs: by examining the variability embedded in the averaged recordings ($\overline{A_N(t)e}$) and estimated
background noise level $\overline{D_N(t)_c}$, the overall outcome of the FFR recordings could be represented using the ratio of these variance values, i.e., the PVR value. It indicates the level of the physiological response that was preserved in the FFR recording. A low PVR value would suggest that the desired signal is likely to be buried in the background noise. On the contrary, a higher PVR value would indicate that the FFR to the voice pitch was present in the averaged recording and it stood out from the background noise.

For the RSL based spectral procedure, it used the known $f_0$ contour of the stimulus as the targeted comparison group and made statistical decisions based solely on the recorded data. Different from the methods previously reported by Jeng, Hu, Dickman, Lin, et al., (2011), the RSL index does not rely on the extracted $f_0$ contour, thus eliminates the need for accurate extraction of the $f_0$ contour from the recording and consequently makes the response detection free of the potential error during the $f_0$ extraction process. Further, during the calculation of the RSL index, the signal and noise regions, where the statistical comparison was performed, were determined by the known $f_0$ contour of the specific stimulus used in the recording. Hence makes the process stimulus-specific: as long as stimulus has a known $f_0$ contour, it could be used in this RSL based spectral procedure and if detected, the elicited physiological recording will contain specific spectral information that only exists in that stimulus.

**Temporal progression and effect of sweeps.** When examining the temporal progression of the PVR values along the temporal axis, although the Mandarin Chinese tokens used in this study have clear and distinctive temporal $f_0$ trajectories (flat, rising, dipping and falling), there were no significant differences between each temporal
segment, for all four tones. This may suggest that although the spectral information in both the stimulus and response could be different in each temporal segment, the response amplitude, as reflected as the PVR obtained in the temporal domain, does not differ that much. This is somewhat contrary with some previous works (Jeng et al., 2010; Krishnan et al., 2004) that some of the tokens that have shaper spectral shape (e.g., Tone 2 and 4, rising and falling) would elicit responses that have more robust pitch-related energy than other tokens (e.g., Tone 1 and 3, flat and dipping). This could be due to the fact that the PVR values provide information solely on the temporal domain and that it would not reflect the subtle difference in the spectral domain. However, depend on the usage of the PVR procedure, such feature will not limit its application, as long as the purpose of the voice pitch elicit FFR test is not to study in detail of the specific spectral information embedded in the recording. And for this purpose, we have the RSL based spectral procedure that is designed for detecting specific spectral information.

Every electrophysiological measurement has to face the dilemma and balance out between the number of sweeps recorded (i.e., the amount of time needed) and the signal-to-noise ratio of the recorded waveforms (Picton et al., 1983; Valdés et al., 1997). In general, to improve the signal-to-noise ratio of the recorded waveform, one would want more recording sweeps so that the noise can be reduced to a minimum in the averaged waveform. But the more sweeps there are, the more time it will need to reach a certain satisfactory level of a certain signal-to-noise ratio (or response detection criterion: $PVR_{critical}$) and that is a major concern in the real world applications of the electrophysiological measurements. In this study, with an appropriate choice of the
stimulus token, the number of recording sweeps needed to reach $PVR_{\text{critical}}$ could be as low as 600 sweeps. This suggests that this automated response detection procedure for voice pitch elicited FFR could accomplish the response detection tasks efficiently without compromising the level of statistical robustness ($PVR_{\text{critical}} = 1.05$, $\alpha = 0.05$). Similarly, for the RSL based spectral procedure, the RSL values obtained from the four stimulated conditions started to divert from those obtained from the control condition from the very first 200 sweeps. Interestingly, the RSL values obtained from the control condition steadily decrease as the number of sweeps increases (0.43 at 200 sweeps and 0.24 at 2000 sweeps). Pending the choice of the $RSL_{\text{critical}}$ from the experimenter, the RSL based spectral procedure should also reach its highest performance point with high efficiency, similar to that of the PVR based temporal procedure. Such efficiency could take the automated procedure to a vantage point when the FFR to voice pitch becomes a useful tool for patients in the clinic.

**PVR based temporal procedure and other temporal statistical procedure for EP.** One major advantage inherent to most statistical procedures is that the results can be compared with a certain critical threshold criterion to determine the presence or absence of a response. This manner of signal processing and data interpretation is simple, robust, and free from potential errors that are related to human judgments. Some of the statistics based response detection procedures could be found in other electrophysiological fields, such as those used in AABR. As mentioned in the review, AABR algorithms use either the Z statistics (Smits et al., 2002) or the point-optimized variance ratio (Sininger et al., 2001). For example, when the Z statistics is used in AABR, the summed polarity
sequences is compared to a specific normative response template and a Z-statistic would be calculated between the sequence and the template. Thus, as the recording sweeps increases, the test statistic would increase until a $z_{\text{max}}$ of 4 is reached, which indicates a “pass” (Smits et al., 2002). For the point-optimized variance ratio method, it utilizes the $F_{sp}$ principle described in the review section, where in general, $F_{sp}$ represents the variance ratio ($F_{sp}$) in the recording. In this method, the recording would continue until an $F_{sp}$ of 3.1 is reached, with the degrees of freedoms of 5 and 250 (Sinninger, 1993). Further exploration of these methods could be found in examples such as the introduction of fixed-multiple-point ($F_{mp}$) method (Smits et al., 2002), where a modified version of the $F_{sp}$ that accounts for a discrete number of noise sources of different power were used to better estimate the SNR in the averaged signal, and the continuous loop averaging deconvolution (CLAD) method (Delgado & Özdamar, 2004), which deconvolves overlapping auditory evoked responses and provides information as to the response characteristics at very high stimulation rates not possible with conventional techniques (Özdamar & Bohórquez, 2006).

Similarly, the PVR based automated procedure takes advantages of the statistical properties of the recorded sweeps. By comparing the averaged response to the estimated background noise, the PVR algorithm acts on each individual recording and output the PVR value, which is used in the automated procedure to compare with the $PVR_{\text{critical}}$ and output a response detection decision. Unlike the Z-statistic based AABR algorithm that requires a response template (Elberling & Don, 1984) to be built for a specific population (e.g., newborn babies) that the researcher is interested in, the proposed automated
procedure does not require any response template to be built. This advantage not only eliminates the need for the user to establish a pre-recorded responses template, but also allows implementation of this procedure on literally all possible patient populations. Because the proposed automated response detection procedure is based on nothing but the statistical properties of the recorded data, this procedure could be utilized for patient groups that differ in age, linguistic background, and pathologies along the auditory pathways. It does not use any pre-stored template, need any manual input, and is self-explanatory and safe from potential human error.

When compared to the point-optimized variance ratio AABR method (Sininger et al., 2000), this automated procedure also makes its decision based on the data itself and the F-statistics, which represents the SNR. However, the ways that the F-statistics are computed are different in these two methods. In the point-optimized variance ratio method, the $F_{sp}$ is calculated by using the variance in the averaged waveform as the numerator ($\text{VAR}(S)$), which represents the signal, and the variance of the sweep to sweep values at a signal fixed time ($\text{VAR}(SP)$), which represents the noise, as the denominator. In our PVR based temporal algorithm, the numerator, i.e., the variance of the signal, is also calculated using the variance in the averaged waveform ($\overline{A_N(t)}$). However, instead of picking a point in the time array, we used the difference between the two sweep banks (odd and even number of sweeps), ($\overline{D_N(t)}$), to represent the noise and to calculate the variance in it. This way, by using ($\overline{D_N(t)}$), the representation of noise in the signal, ($\overline{n(t)}$), as in Formula 5, is a direct and more accurate estimation of the real non-
physiological activities, i.e., the noise, in the EEG recording because of the way it is derived.

**RSL based spectral procedure and other spectral analytical procedures for EP.** Analytical procedures for EP in the spectral domain are nothing new. These analytical procedures have employed a variety of commonly used methods such as band-pass filtering and Fourier analysis. For time variant signals, such as those used in ASSR and voice pitch elicited FFR, the applications of methods such as FFT, crosscorrelation and autocorrelations are widely seen in the literatures (John & Picton, 2000b, 2000a; Özdamar & Bohórquez, 2006; Skoe & Kraus, 2010; Valdés et al., 1997).

Different from the reported by Jeng, Hu, Dickman, Lin, et al., (2011b), which uses solely the information from the spectral-domain and searches for the maximal spectral density in a predetermined frequency range, the new statistics-amended spectral algorithm utilize both information from the spectral energy density distribution and the statistic property of such distribution. First off, the way that the new statistics-amended spectral algorithm extracts the f0 contours differs from the old one, which extracts the f0 contour by itself. This algorithm, however, uses the information embedded in the data to determine the stimulus that was used in the recording session. With the stimulus (one of the four Mandarin syllables) information, the algorithm first read the pre-stored f0 contour array of the stimulus. Similar to the “narrow-band spectrogram” algorithm, the new algorithm starts searching for the maximal spectral density value in each windowed segment. However, instead of doing so in a fixed pre-defined frequency range, it will extend its searching range based upon the frequency value of the stimulus’ f0 in that
segment. In other words, the searching ranges changes dynamically with the temporal advancement of the recorded waveform.

Such difference gives the current spectral procedure an edge over the older one in that it relies on the information that is based on the f0 of the stimulus. While the older procedure had to extract the recording f0 itself and then carry on with the computation, this procedure uses the stimulus f0 information to make decisions upon. Its response detection technique focuses on the specific spectral information of each specific stimulus. Thus, when different stimuli were used, e.g., the four Mandarin tones, the responses could be interpreted as the physiological activities that are specific to each stimulus.

Response detection performance. When compared with human judgment and reviewing the corresponding ROC space and curves, both the PVR based temporal procedure and the RSL based spectral procedure revealed relatively promising performances. For the PVR based procedure, different tones gave results that at various levels, when PVRcritical is set at 1.05: Tone 3 provided a sensitivity score at 86.21% and a perfect specificity score at 100%, suggesting its possible application as the primary token to be used in the PVR based procedure. Tone 2 had a moderate 62.96% and 66.67% scores for sensitivity and specificity, respectively, suggesting that it could also be potentially used in the PVR procedure. Tone 1 had a higher specificity of 75% while its sensitivity score was found at 42.31%. Finally, Tone 4’s 54.55% and 50% scores suggests that these two tones might not be the first choices to be used in the PVR based temporal procedure, when the other two tones were readily available. Among the four Mandarin tones, Tone 3 (dipping) required the least amount of sweeps (600) to reach the
PVR_{critical} value. It also provided the highest sensitivity (86.21%) and specificity (100%) when the automated procedure outputs were compared to subjective expert determination. This finding is interesting in that previous studies on Mandarin tone elicited FFRs (Jeng et al., 2010; Krishnan et al., 2004, 2005;) reported that when the FFRs were recorded from native Mandarin speakers, it was the tones that had mono-directional pitch contours (Tone 2, rising and Tone 4 falling) that resulted in relatively larger FFR amplitude and more accurate pitch-tracking ability, as opposed to tones with bi-directional pitch contour such as Tone 3. The superior performance of Tone 3 observed in this study could be contributed to, at least in part, that Tone 3 used had a slightly lower fundamental frequency than the other three tones used in this study. It is possible that the neural responses elicited in each single sweep may not be perfectly aligned with the neural responses elicited in the next sweep. This slight misalignment between the recorded sweeps may affect the performance of this automated procedure, because the PVR algorithm was assigning recording sweeps into two separate banks constantly. As a result, such slight phase delay between the two recording banks could favor a stimulus tone that has a lower fundamental frequency. Because Tone 3 performed the best in this automated procedure, it can be a viable stimulus when this PVR based procedure is put in use for future applications. Here, notice that these results were based on the PVR_{critical} value that was derived from degrees of freedoms of 4999 and 4999. If different degrees of freedoms were utilized in a certain way, which could be found in later discussions, the result would be different (Sininger et al., 2001).
On the other hand, the RSL based spectral procedure provided ROC curves that were more favorable. As indicated in Table 5, all four tones showed sensitivity that could still hold up until higher acceptance rate (RSL_{critical}). Among the four tones, the specificity of Tone 1 provided a TPR at 93% even the RSL_{critical} were as strict as 90% (i.e., the procedure will only consider a recording a voice pitch elicited FFR under the condition that at least 90% of the windowed segments have significant spectral energy that stands out from the noise). TPR score of 90% were seen when Tone 2 had an RSL_{critical} of 85% and Tone 3 had the RSL_{critical} at 80 %. These results suggest that as a response detection procedure, the RSL based spectral procedure could detect voice pitch elicited FFR responses even with acceptance rate (i.e., RSL_{critical}) that are very strict correctly. Similarly, when it comes to the specificity, as indicated in Table 6, all four tones also showed FPRs (i.e., 1-specificity) that could still remain relatively low at a higher acceptance rate (RSL_{critical}). Also similar to the TPR data, the FPR provided by Tone 1 was recorded at zero at 70%. Tone 2 provided performance similar to that of Tone 1. Tone 3 and Tone 4 reached zero FPR at RSL_{critical} of 80% and 100%, respectively. Interestingly, all four tones had single digit FPRs at only 65% RSL_{critical} level, indicating that when no matter which tone were to be used in the RSL procedure, the probability that the response detection procedure would mark a non-response recording as an FFR response (i.e., the Type I error) is below 10%, when a much moderate RSL_{critical} was chosen. If the RSL_{critical} were to be set at 80%, such probability to make a Type I error would further reduces to zero, except Tone 4.
From the statistics and disease epidemiology point of view, the ROC performances of the two procedures have values that can be seen in the immediate reality (Jeng at al., 2010). A sensitive test takes preference whenever the probability of FPR is high or whenever it is needed to reduce the probability of errors (Altman & Bland, 1994). In other words, sensitivity (i.e., power) is primarily used to rule-out the existence of a disease. On the other hand, specificity (i.e., 1-FPR) is often used to confirm an existing symptom or diagnostic impression; highly specific tests indicate lower FPR. Thus, for any automated procedure to be used in the clinical settings, the advantages and disadvantages of the algorithm should be carefully weighted and balanced. For example, when treating diseases that may not introduce serious consequences, it would be appropriate to use a test that yields the larger power and an acceptable false-positive rate, while a strict criteria or higher FPR is absolutely need when screening for diseases that carry serious consequences if misdiagnosed or not treated early, such as tumors and severe-profound hearing loss at birth.

For our automated voice elicited FFR response detection procedure, the implantation of the above discussed principle makes the procedure very versatile and yet provide favorable performances at both ends. For example, in the PVR based temporal procedure, Tone 3 provided a TPR score of 86.21% and a FPR of zero, meaning that if it were used as a stimulus token and if a participant could successfully encode the voice pitch information, there is a 86.21% probability that our procedure could detect the such ability that is embedded in the EEG, as reflected by his or her FFR. On the other hand, if the purpose of the test were to confirm certain symptoms, e.g., the lack of pitch tracking
ability, the zero FPR of Tone 3 in the PVR procedure would excel in that scenario in that if the patient indeed lacks the ability, there was zero chance that the procedure could miss such deficit. This kind of the high specificity (i.e., low FPR) makes the procedure ideal when an effective screening tool is needed. For the RSL based spectral procedure, the usage of it could be more versatile as the RSL_{critical} (acceptance rate) were selected. For example, if higher power is needed depending on the specific clinical setting, an RSL_{critical} can be set at a stricter 85% for Tone 1 and more forgiving 65% for Tone 2 and 3. But if a low FPR is needed in situations such as screening purposes, to make sure there is no missing “impaired” recording, the RSL_{critical} of Tone 1and 2 needs to set as at least 70% percent, while a higher 80% is needed for Tone 3. Further, if higher power (sensitivity) and lower FPR are needed at the same time, which means a very versatile yet high demanding setting, Tone 1 could have very favorable performance when RSL_{critical} is set at 70-85%, which would give a 100% power and zero FPR. If the demanding were not to such extreme, Tone 2 and 3 could also give promising result at 65 or 70% RSL_{critical} level, where they could provide near perfect (97%) power and single digit FPR numbers (3 or 6%).

**PVR based temporal procedure and RSL based spectral procedure.** As mentioned previously, the PVR based temporal procedure and RSL based spectral procedure could both be considered powerful and yet versatile tools to automatically detect the voice pitch elicited FFR. These two procedures share an important common ground that the essential part of their detection algorithm are both based on statistical decisions of the data itself, independent of any manual input or pre-existed template. Yet,
they do have differences that make them excel in different domains and settings. First off, they examine the FFR data in two fundamentally different domains, one in the time domain (PVR procedure), and the other in the spectral domain (RSL domain). Further, the spectral specificity of this RSL based spectral procedure is also a major difference with the above PVR temporal procedure, where the detection of a response is made upon the temporal property of the recordings. While the PVR procedure focuses on the overall quality of the EEG recording and makes response detection decision on such quality, the RSL based spectral procedure focuses on the spectral property in the recording that is specific to the current stimulus. Such differences can come in handy when we were to examine a participant’s response to voice pitch using FFR in different scenarios. For example, if voice pitch elicited FFR were to used primarily as a screening tool for detecting certain audiological syndromes, such the degraded pitch resolving ability in children with ASD, we could use the PVR based temporal procedure to test the overall quality of the response. On the other hand, if the voice pitch elicited FFR were to be used when certain aspect of the spectral information needs to be examined in detail, e.g., when different music notes or language tones were used with an attempt to elicit different responses in participants, the RSL based spectral procedure could see its application.

**Limitations and Future Considerations**

Several factors that might limit the efficiency and portability of this automated response detection procedure. First, the heavy computational demands (including separation of the two banks, addition, subtraction, F-statistics, and comparison with response threshold detection), may be a limitation factor, depending on how often the
PVR values are calculated throughout the time course of data collection. The more often the computational steps are iterated, the heavier the demand it will be, which might potentially slow down the performance of a computer and the data collection process. Secondly, the duration of each recorded sweep determines the number of data points per sweep, which in turn determines the degree of freedom in the statistical computation of the automated response detection procedure. One possible alternative solution would be to lower the sampling rate of the recording so that the number of data points collected for each individual sweep could be reduced. This solution, however, has a risk of losing some critical morphological details of the response and will reduce the degree of freedom in the statistical analysis. Specifically, lowering the recording sampling rate would elevate the PVR_{critical} value and will make the recording even longer to reach the elevated response detection threshold. Lastly, note the choice of the PVR_{critical} value will vary depending on the specific α level set in the statistical computation. Further research focuses on examining the effects of these limiting parameters would shed light on the FFR applications for basic scientific research and clinical applications.

**Clinical Applications**

Traditionally, to reliably interpret the ABR waveforms, one would need to be an experienced clinician or researcher who has substantial training and experience in the field. By eliminating the observer’s input in the decision making process, this automated procedure is independent from the observer’s subjective interpretation and thus it would require less training for an end user to record and determine the presence or absence of an
FFR to voice pitch. This advantage promotes the efficiency and portability of the proposed technology and thus reduces the cost in its research and clinical applications.

The main need for detection and assessment of an electrophysiological measurement is ultimately related to clinical evaluation of auditory stimulation in humans. Applications such as hearing screenings and auditory evaluation pertain particularly to infants, children, and other individuals from whom reliable behavioral audiometric information cannot be obtained. For analyzing these measurements, extraction of reliable response parameter is essential for assessing the effects of ontological diseases in patients.
References


