uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

u Ottawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Marilyn Laroche

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Biomedical Engineering)

GRADE / DEGREE

Faculty of Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

A Study of Auditory Speech Processing using Brainstem Evoked Responses under Quiet and Noisy Conditions

TITRE DE LA THÈSE / TITLE OF THESIS

Hilmi Dajani

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Yuu Ono

Christian Giguère

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

A Study of Auditory Speech Processing using Brainstem Evoked Responses under Quiet and Noisy Conditions

by

Marilyn Laroche

A thesis submitted to the Faculty of Graduate and Postgraduate Studies in partial fulfillment of the requirements for the degree of

Master of Applied Sciences

The Ottawa-Carleton Institute for
Biomedical Engineering

School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario
2010

Your file Votre référence
ISBN: 978-0-494-74219-8
Our file Notre référence
ISBN: 978-0-494-74219-8

Abstract

This work investigates auditory speech processing through measurement of brainstem responses to the synthetic vowel /a/, and variants containing the first formant (F_1) only or the second and third formants only, in quiet and in white background noise. The results show that the response at the fundamental frequency (F_0) is more robust to noise than the response at F_1. This study also provides an insight into the potential locations where the pitch of resolved and unresolved harmonics is extracted in the auditory system, and how the frequency regions dominated by resolved/unresolved harmonics contribute to the response at F_0. Based on the results, a conceptual model of the neural processing underlying the generation of the electrophysiological response at F_0 is proposed.

Acknowledgments

I would like to take this opportunity to thank my thesis supervisor Prof. Hilmi Dajani for his direction, strong support, and prompt feedback without which this work would not be possible.

I would also like to thank Prof. André Marcoux from the Faculty of Health Sciences at the University of Ottawa for his support and advices in this research.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ABR	Auditory Brainstem Response
Allformants	Stimulus variant containing F_1, F_2 and F_3
AM	Amplitude modulated
ANOVA	Analysis of Variance
ASSR	Auditory Steady-State Response
CF	Characteristic Frequency
CN	Cochlear Nucleus
dB SPL	Decibels (Sound Pressure Level)
dB HL	Decibels (Hearing Level)
EEG	Electroencephalogram
EFR	Envelope Frequency Response
EP	Evoked Potential
FFR	Frequency-following Response
fMRI	Functional Magnetic Resonance Imaging
F_c	Carrier frequency
F_m	Modulation frequency
F_1Only	Stimulus variant containing the first formant only
F_2&F_3Only	Stimulus variant containing the second and third formants only
F_0	Fundamental (or Pitch) Frequency
F_1	First formant
F_2	Second formant
F_3	Third formant
F_4	Fourth formant
ICC	Inferior colliculus
LL	Lateral Lemniscus
MGB	Medial geniculate body
rms-SNR	Root-mean-square Signal-to-noise ratio
SLM	Sound Level Meter
SNR	Signal-to-noise ratio
SpEPs	Speech Evoked Potentials
VA	VA onset complex (transient response)

Chapter 1 : Introduction

1.1 Motivation

It is hard to imagine how one could communicate with others without using speech and hearing. Many children and parents are facing this issue every day. Improvements in the diagnostic techniques for hearing impairment could allow customizing the rehabilitation of children and adults with hearing disabilities to their specific needs. Unfortunately, we are still at the early stages of understanding the physiological events underlying the processing of speech in the auditory system.

Over the years, the scalp-recorded auditory brainstem response (ABR) has allowed researchers and clinicians to obtain a better understanding of the structures and neural mechanisms that define the human auditory processing. One can use the ABR to detect the presence of lesions or malformations in the auditory system, or to evaluate the hearing threshold of a subject suspected to have hearing impairment. Collecting the evoked responses to a speech stimulus (speech ABR) is a recently developed non-invasive technique which provides important information about auditory processing of speech up to the brainstem level. Several studies have demonstrated that using speech as an acoustic stimulus can be a beneficial diagnostic tool. With the speech ABR, clinicians can assess the condition of the peripheral and central auditory neural pathways of children who have language-based learning problems or central auditory processing disorders (Greenberg and Ainsworth, 2004; Russo et al., 2004; Johnson et al., 2005; Kraus and Nicol, 2005; Burkard et al., 2007; Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010).

Using speech as an acoustic stimulus to develop new and more efficient clinical tests is a must. It provides an additional value to the rehabilitation process of children with hearing disabilities over the most common currently used audiological tests that employ pure tones, clicks, and other artificial signals. Speech-evoked ABR may also complement other existing tests that have been conducted for evaluating speech intelligibility with hearing aid usage (Johnson et al., 2005; Aiken and Picton, 2006; Johnson et al., 2008). The timing of the major activities in the speech-evoked ABR provides indications about the quality of neural

encoding of the speech stimulus in the auditory system (Johnson et al., 2005; Johnson et al., 2008). Previous work has been done with both natural and synthesized speech stimuli. The use of a synthesized stimulus allows more control over its content and simplifies the analysis of the collected responses by eliminating variations in the spectral components of the signal. Moreover, because of the fixed harmonic components in a synthesized stimulus, it allows a faster detection of the main features in the evoked response.

Speech is an acoustic signal that originates in the vibrations of the vocal folds of the speaker and is modified by the articulators in the oral cavity (Palmer and Shamma, 2004). For the listener, the speech signal is shaped by the frequency response of the outer and middle ear. The basilar membrane, the main component of the cochlea located inside the inner ear, can be modelled as a spectral analyzer which allows the decomposition of the speech signal into its frequency components before being further processed by the central auditory system (Assmann and Summerfield, 2004). The central auditory neural mechanisms involve both temporal and place encoding of the acoustic cues forming the speech signal, but are not well understood.

The neural activity in response to speech propagates to the scalp and is detected as an evoked response. The resulting speech-evoked ABR is made of two main components: the transient and the steady-state response. Extensive research has been conducted using various acoustic stimuli, including speech, to identify the origins of the main neural events characterizing the transient response. Less is known about the neural mechanisms associated with the steady-state response. There is still no consensus about their origin in the central auditory system, although they are believed to originate in the upper brainstem when a so-called "vertical" electrode montage is used. Moreover, it is not known how different types of noise affect the components of the response.

Background noise has an important impact on everyday communications, especially for the hearing impaired and children with language based learning disabilities (Anderson et al., 2010). Hence, this suggests the importance of taking into account the role of noise during the rehabilitation of hearing impaired subjects (Pittman and Wiley, 2001). The processing of the

acoustic–phonetic features of speech sounds relevant to speech perception begins no later than 5–7 ms from the time the auditory signal enters the ear (Krishnan and Gandour, 2009). Decoding the phonetic content of a message in such a short period of time in a noisy environment requires complex auditory processes. Since we are rarely exposed to situations comparable to a controlled quiet environment, such as the one in an acoustically and electrically shielded booth, studying the effect of noisy speech on the auditory system is at the heart of the study of speech communication. Therefore, adding noise to a speech stimulus would contribute to a better understanding of speech evoked potentials, and potentially lead to improvements in the diagnosis of the hearing-impaired and their rehabilitation through new designs of hearing aids. Once we have a better understanding of the processing of speech in the auditory system, developing methods and techniques to reduce the background noise and improving the detection of speech stimuli would follow. For now, since we know how noise is usually detrimental to speech perception, we must concentrate on understanding what noise actually does to the central auditory processes involved in normal communication.

1.2 Objectives

The purpose of the present work is to investigate the physiological events underlying speech processing in the auditory nervous system at the brainstem level, including in background noise.

The objectives can be summarized as follows:

- To study the noise robustness of the brainstem response at the fundamental frequency F_0 and at the first formant F_1.
- To evaluate how the resolved/unresolved harmonics regions contribute to the response at F_0.
- To investigate the origin of the response at F_0 to resolved and unresolved harmonics.

The first part of the experiment is intended to study the robustness of the brainstem response to noise, by looking at how the system is susceptible to variation in signal-to-noise ratio (SNR) using the synthesized vowel /a/ as stimulus. This stage includes performing statistical analysis on the response at the fundamental frequency (F_0) and at the first formant (F_1).

The second part of the experiment aims at studying the contribution of two regions of the frequency spectrum to the response at the fundamental frequency (F_0). The first region is dominated by resolved harmonics, and the second is dominated by unresolved harmonics. We are also looking at the response at F_0 generated by the stimulus which contains both regions. The analysis of both parts of the experiment involves looking at how the amplitude and SNR of the responses are affected by the signal content, and is done by performing (repeated measures) two-way ANOVA statistical analysis followed by post-hoc tests.

We are also interested in investigating whether or not the auditory system processes evoked-potentials linearly, particularly for the response at F_0. This is done by comparing the Vector Sum of the responses to resolved and unresolved harmonics with the response vector due to the stimulus containing both frequency regions. To investigate the origin of the response at F_0, we also measure the latency between the stimuli signal and the evoked response for the quiet and noisy conditions.

1.3 Thesis Outline

This work is organized in six chapters. The remaining five chapters are briefly described here.

Chapter 2: This chapter provides an up-to-date review of what is known about the auditory nervous system and its neural mechanisms. A brief overview of the structure of the auditory system is presented along with the major sound processing concepts (pitch and pitch processing, place theory, temporal coding theory, and phase-locking). A presentation of auditory evoked potentials with specific coverage of speech-evoked potentials at the brainstem and cortical level is included. A description of the characteristics of vowel sounds is also included along with examples. Finally, a summary of the commonly-used methods for synthesizing speech sounds wraps up this chapter.

Chapter 3: This chapter presents the methodology of the research. It provides a detailed description of the participating subjects, and the methods employed in generating and collecting the speech ABRs. The experimental protocol is also included along with its main objectives. A detailed description of the construction of the stimuli follows, providing necessary information about its calibration, along with the tests performed to verify the absence of electromagnetic leakage. A brief description of coherent averaging and its importance in the measurement of ABRs is included as well. The steps that were done to verify the quality of the recorded signal and minimization of the power-line noise contamination in the responses are also presented. Finally, the statistical analysis and set-up of the ANOVA follows.

Chapter 4: This chapter presents the results of the analysis performed on the measured evoked responses. The effect of noise on the transient and sustained responses, and the impact of the stimulus content on the sustained responses are presented. It also includes the analysis of the linearity of the response of the auditory system at the brainstem level at F_0, and estimates of the latency of the response.

Chapter 5: This chapter provides in-depth discussion of the results. It summarizes important information about the effect of noise on the transient and sustained responses with respect to previous work. Next, a section describing how the auditory system (peripheral, brainstem, cortical) processes resolved and unresolved harmonics is included. This chapter also includes a section describing the possible origins of responses to resolved/unresolved harmonics in the auditory system.

Chapter 6: This chapter concludes the thesis with a summary of the major findings and contributions. Finally, suggestions for future work on speech-evoked ABR are also included.

Chapter 2 : Background

2.1 The Auditory Nervous System

As Møller describes in his reference book: *Hearing: Anatomy, physiology, and disorders of the auditory system*, the auditory nervous system is "the most complex of all sensory pathways" (Møller, 2006b). Researchers have not yet discovered all the mysteries behind the properties of this network of nerves and more work is to be done. This section will describe the principles and main function of the auditory nervous system.

First of all, the auditory nervous system is divided into the ascending and descending pathways. The ascending pathways have two subdivisions: the classical and non-classical pathways. The ascending and the descending pathways are differentiated in terms of the direction of the signal; that is, the ascending pathways process and relay auditory information from the organ of Corti to the temporal lobes of the auditory cortex whereas the descending pathways are from the temporal lobes to the organ of Corti (Møller, 2003; Møller, 2006a).

2.1.1 The Ascending Pathways

The classical ascending auditory system

In this pathway, there are three main relay nuclei located between the auditory nerve and the primary auditory cortex: the cochlear nucleus (CN), the central nucleus of the contralateral inferior colliculus (ICC) and medial geniculate body (MGB) (Møller, 2003; Møller, 2006a). There are two partly independent ascending auditory pathways, starting from each ear, with connections present at various levels between the two hemispheres. A simplified classical ascending auditory pathway schematic diagram is shown in figure 2-1.

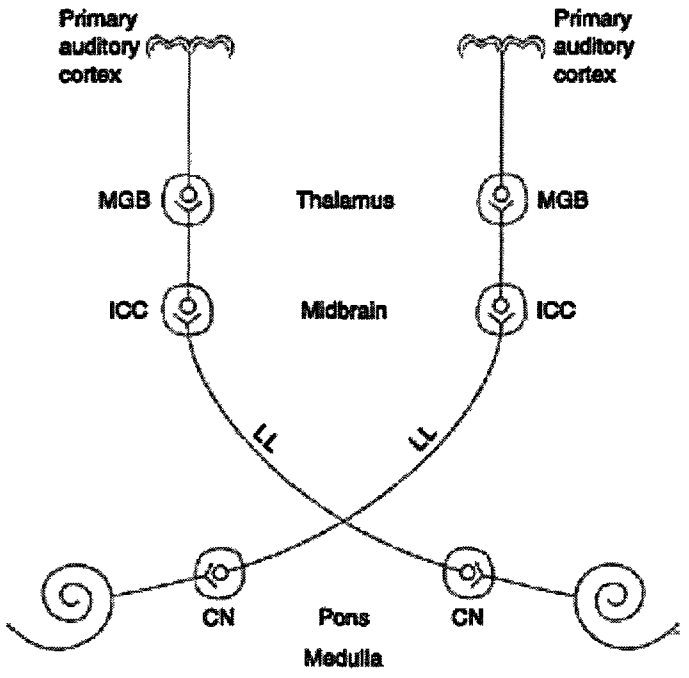

Figure 2-1:Simplified classical ascending auditory pathways schematic [reprinted with permission from (Møller, 2006a). Copyright 2006, Elsevier.]

Located in the lower brainstem, the cochlear nucleus is the first relay on the ascending auditory pathway. The inferior colliculus (ICC) is the second major relay after the CN. Located in the midbrain, the ICC channels all the auditory information coming from the lateral lemniscus (LL). Since human beings have two ears, there are two partly independent ascending auditory pathways where connections exist at various levels between the two hemispheres. The Medial Geniculate Body (MGB) relays all the auditory information from the ICC to the auditory cortex. Numerous connections exist between the primary auditory cortex and other regions of the brain including the opposite hemisphere of the brain (Møller, 2003; Møller, 2006a). Connections existing between the auditory cortex and the rest of the brain are not predictable; they can actually vary due to neural plasticity (Møller, 2006a). This is an interesting aspect that several researchers have considered when studying whether the auditory system can "re-configure" itself to better process speech, especially for people having a damaged or underdeveloped nervous system (Kraus et al., 1995b). Beyond the auditory cortex, other parts of the brain, such as the Wernicke's area, may have an indirect role in speech signal processing (Harpaz et al., 2009). However, because cortical neurons have difficulty following the rapidly changing temporal patterns of speech, there is evidence

that most of the auditory speech signal processing is performed in the subcortical structures rather than at or beyond the auditory cortex (Griffiths et al., 2001; Bartlett and Wang, 2007).

Non-classical ascending auditory system

Non-classical ascending auditory pathways refer to pathways which are not as clearly defined as the classical pathways (Møller, 2006a). They receive signals from other sensory systems such as the sense of touch as input. Since these pathways are not directly related to the question dealt with in this thesis, they will not be considered further.

2.1.2 Descending Pathways

The descending efferent pathways can be classified into the corticofugal system and the olivocochlear system (Møller, 2003; Møller, 2006a). They are seen as reciprocal to the ascending pathways as the two pathways run parallel to each other. Studies suggest that the efferent system may play a role in speech encoding resulting from language or musical experience (Krishnan et al., 2005; Xu et al., 2006; Parbery-Clark et al., 2009). A refinement of the neural mechanisms or plasticity in the auditory system specific to the efferent system might produce an enhancement or sharpening of particular acoustic features from the speech stimulus. This could improve pitch extraction (F_0) and robustness to noise.

2.2 Acoustics of Vowels

The acoustics of the speech signal can be understood in terms of the influential *source-filter model of speech production* (Johnson et al., 2005; Kraus and Nicol, 2005). This model identifies the source of speech as the vibration of the vocal folds, which determines the fundamental (or pitch) frequency F_0, whereas the filter relates to thetransfer function of the oral cavity that changes during articulation. The resonant peaks of the filter determine the formants frequencies. In this model, it is commonly assumed that there are no interaction between the source and the filter.

Vowels are preferred as speech stimuli because they evoke more robust responses than consonants, which are rich in acoustic information but are short in duration and low in spectral energy making them vulnerable to background noise. The fundamental frequency F_0

of vowels depends on the gender of the speaker. The fundamental frequency for female speakers is the range of ≈175-300 Hz, while for males it is in the range ≈75-175 Hz (Greenberg and Ainsworth, 2004). F_0 also depends on the age and the emotional state of the speaker. The phonetic information in vowels is carried by the *formants*, which are associated with filter characteristics of the vocal cavity. The identification of vowels is generally done by analyzing the frequency of the first two formants (Peterson and Barney, 1952).

Figures 2-2 and 2-3 depict the Fourier amplitude spectra of a 300 ms-synthesized-speech vowel stimulus /u/ as in 'hood' and /a/ as in 'hod'. Each stimulus contains three formants that correspond to the peaks of the spectral envelope of the harmonics of the fundamental frequency F_0. In both cases, F_0 is 100 Hz, corresponding to a male speaker. For the case of the vowel /u/, we have F_1: 300 Hz, F_2: 870 Hz and F_3: 2240 Hz. For the case of the vowel /a/, the first three formants are different (F_1: 700 Hz, F_2: 1220 Hz and F_3: 2600 Hz). Also shown in figure 2-4 is the spectrum of /a/ with added stationary white Gaussian noise at an SNR of -5 dB.

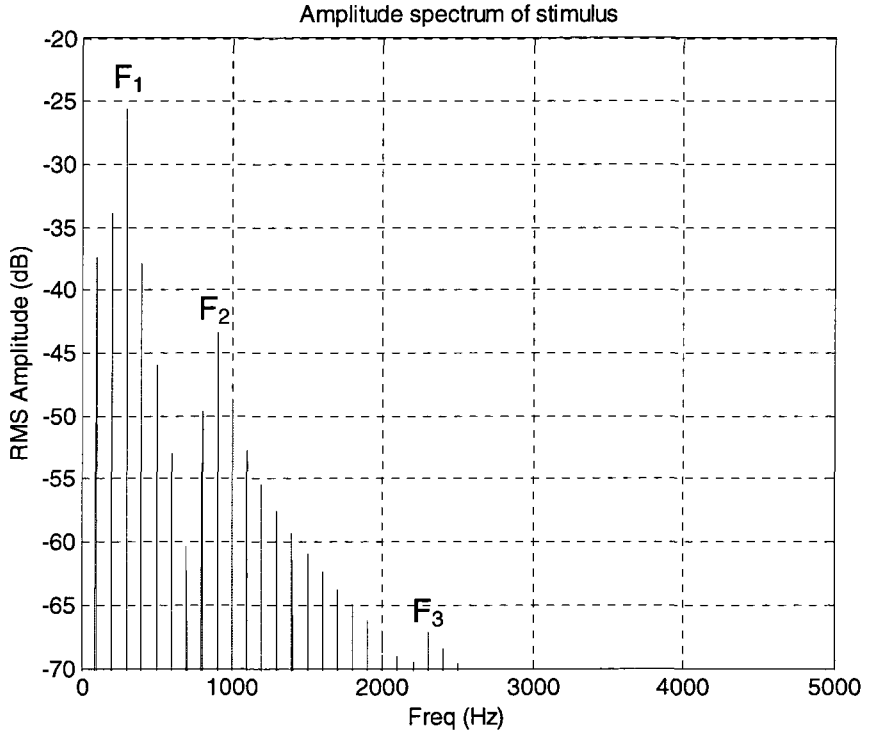

Figure 2-2: Spectrum of /u/ with the first three formants and the harmonics of the fundamental

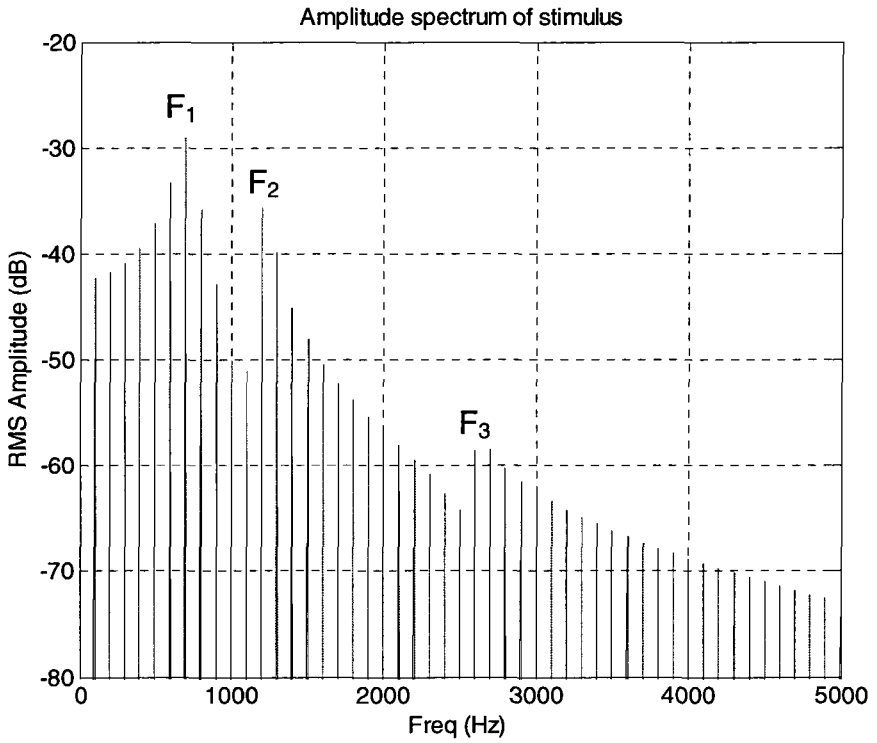

Figure 2-3: Spectrum of vowel /a/ with the first three formants and the harmonics of the fundamental

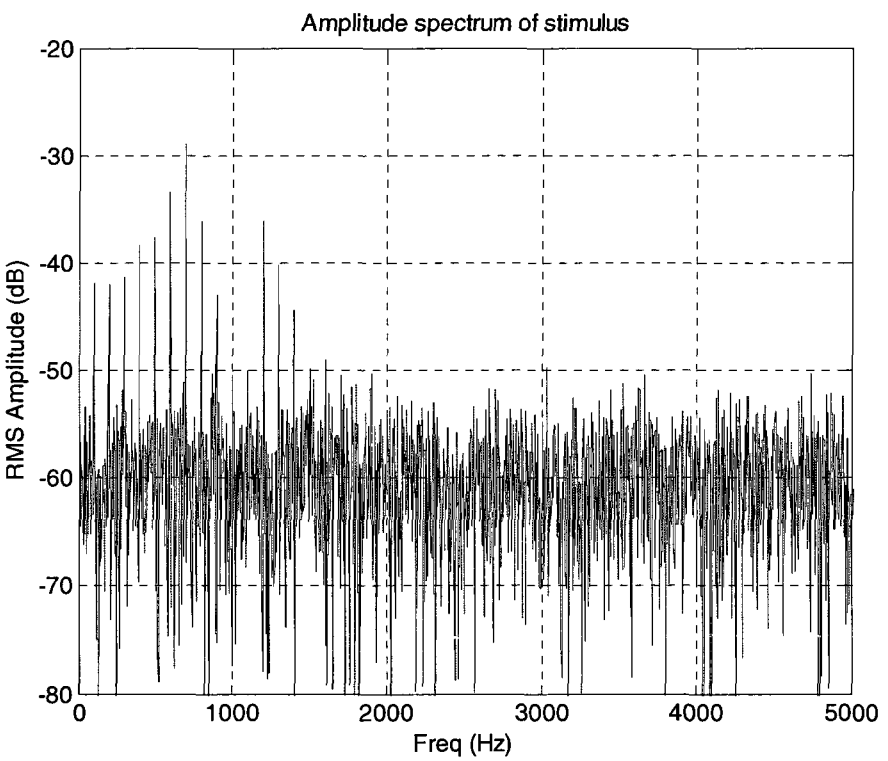

Figure 2-4: Spectrum of /a/ with added stationary white Gaussian noise at an SNR of -5dB

2.3 Phase-locking, Place and Temporal Coding Theories

The frequency content of an acoustic stimulus is encoded in the inner ear (cochlea) from the movement of hair cells that correspond to specific regions of the basilar membrane. If we look at the structure of the basilar membrane, we see that the proximal regions are more sensitive to higher frequencies (>10 kHz) than the distal regions where sensitivity is the highest for lower frequencies (<1 kHz) (Møller, 2006b). Figure 2-5 depicts the structure of the basilar membrane along with the locations where the neural response is the highest for each frequency.

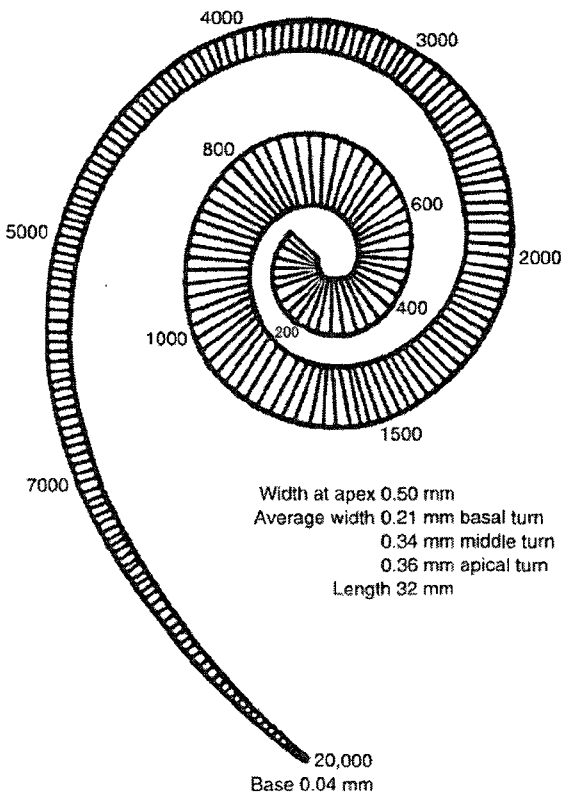

Figure 2-5: Basilar membrane frequency selectivity [reprinted with permission from (Møller, 2006b). Copyright 2006, Elsevier.]

Under the best of conditions, the human ear can detect speech and non-speech sounds between 20 Hz and 20 kHz (Møller, 2006b). When stimulated by an acoustic signal, the auditory nerve activity becomes phase-locked to certain components of the stimulus. Generally, this neural activity synchronizes to the portions of the frequency spectrum that are

of greatest magnitude (Greenberg and Ainsworth, 2004). Phase-locking means that the auditory nerve fibres fire at a specific phase angle of a sound wave period during steady-state and periodic portions of the sound wave. If we could imagine the representation of a sinusoidal sound wave in the time domain traveling through the auditory system, we would notice groups of neurons that would be firing at a specific phase of the sound wave but not necessarily in every cycle (see figure 2-6).

Phase-locking degrades as the frequency increases. Generally, for frequencies higher than 4 kHz phase-locking will be poor to inexistent in the auditory nerve (Greenberg and Ainsworth, 2004; Bear et al., 2007). Above this frequency, the auditory system is thought to rely on tonotopic maps (Bear et al., 2007). Tonotopy, also called the *Place Coding theory*, refers to the nerve cells being "anatomically organized" according to their tuned frequency throughout the ascending classical pathway (Culler et al., 1943; Møller, 2006a; Goldstein, 2009). Unfortunately, the Place theory has its limitations. For instance, this theory alone cannot explain why we can still perceive a pitch of 100 Hz when the 100 Hz signal is not physically present in the stimulus (Gelfand, 2001).

Another theory which describes the processing of sound in the auditory system is the *Temporal Coding theory*. This theory suggests that each auditory nerve fibre generates neural firing that reflects the temporal pattern of the sound wave (Gelfand, 2001; Møller, 2003; Møller, 2006b). The neural firing pattern is transmitted via the ascending auditory pathways and is analyzed by various nuclei up to the auditory cortex.

The nerve fibres located inside the cochlear nucleus and in the auditory nerve phase-lock strongly to frequencies below 1 kHz (Rhode and Greenberg, 1994; Assmann and Summerfield, 2004; Greenberg et al., 2004b). At the auditory cortex level only phase-locking to slow modulations (<10 Hz) will be present (Assmann and Summerfield, 2004). Therefore, for higher portions of the ascending pathway, phase-locking becomes poorer for high frequencies. Even if phase-locking is not strong in the higher regions of the ascending auditory pathway, the encoding of pitch is still done throughout the auditory system up to the

auditory cortex possibly using place coding (Griffiths et al., 2001; Bartlett and Wang, 2007; Krishnan and Gandour, 2009).

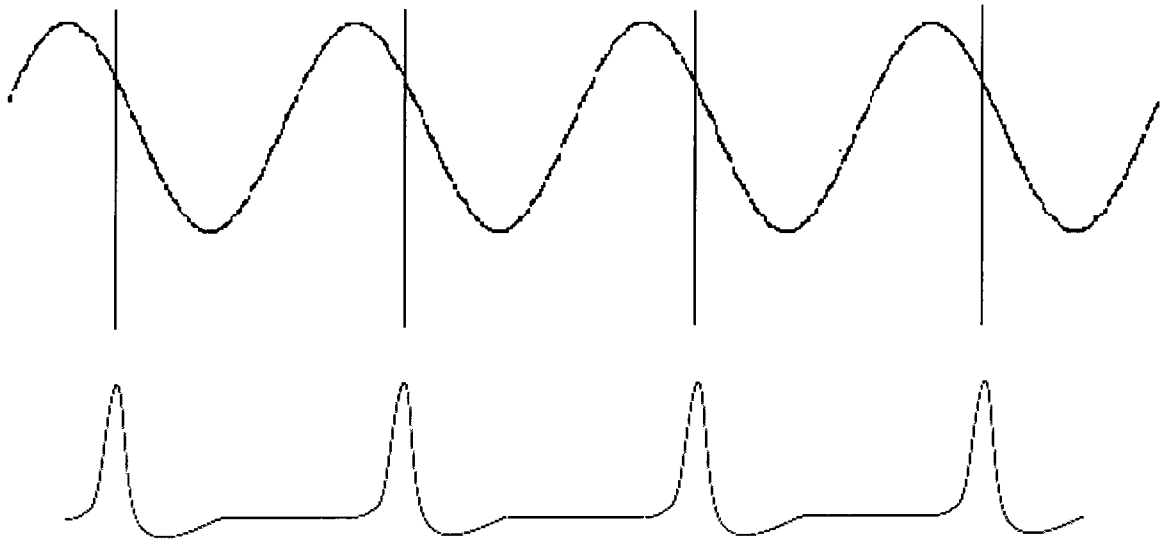

Figure 2-6: Representation of a sinusoidal sound wave in time domain and the generation of action potentials by a phase-locked neuron at every cycle

Phase-locking is important because it allows the temporal representation of sound frequency and so could form the basis of pitch processing in the auditory system. In addition, phase-locking generally follows the first and the second formants of the speech signal if they are located under 1 kHz. If background noise is present in the speech stimulus with most formants still rising above the noise level, the hair cells would still be able to phase-lock to those specific frequencies (Assmann and Summerfield, 2004). Noise in an acoustic stimulus such as speech creates irregular patterns of nerve activity, since there is no temporal structure associated with it (Sachs and Young, 1979; Greenberg and Ainsworth, 2004; Johnson et al., 2005). The randomness created by noise makes it difficult for the hair cells to synchronize to the stimulus. However, with a minimum presence of regular patterns in the firing activity of some of the auditory nerve fibres, the auditory system could probably still extract the dominant frequency components of the sound. For example, having this synchronization from the auditory periphery and brainstem to F_0 has been shown to improve the effective signal-to-noise ratio of the neural response by 10 to 15 dB (Rose et al., 1967; Greenberg and Ainsworth, 2004).

2.4 Resolved and Unresolved Harmonics

Peripheral auditory processing can be thought to include a series of bandpass filters with increasing center frequencies and bandwidths that increase correspondingly. For a complex signal containing multiple harmonics such as speech, the low-ordered harmonics of the fundamental's are considered to be 'resolved' because they are processed *individually* by the narrowband filters. The outputs of these filters are therefore at the frequency that corresponds to the closest harmonic of the speech signal. At higher frequencies, multiple harmonics of the fundamental are combined at the output of the wider bandpass filters, and are hence called 'unresolved'. With unresolved harmonics, the outputs of the bandpass filters are periodic at the fundamental frequency of speech. To understand this, consider the case where two harmonics are combined at the output of a bandpass filter. This results in an AM signal having the 'difference-in-frequency-between-the-harmonics' being the modulation frequency. The boundary between the resolved and unresolved harmonic regions depends on the fundamental frequency (F_0). When F_0 is located in the 100-200 Hz region, as is the case for the stimulus used in this study, high-order harmonics above the 10^{th} are believed to correspond to unresolved harmonics (Houtsma and Smurzynski, 1990; Micheyl and Oxenham, 2004).

Figure 2-7 was taken from a study investigating the temporal pattern of the neural activity of peripheral auditory fibers in cats in response to the synthetic speech stimulus /da/ (Secker-Walker and Campbell, 1990). The neurogram represents neural activity in fibers of different characteristic frequencies (CF) arranged tonotopically along the y-axis. It can be observed that the response at the fundamental periodicity (of approximately 8 ms) is quite strong for both lower CF fibers that presumably respond to resolved harmonics, and for higher CF fibers that likely respond to unresolved harmonics.

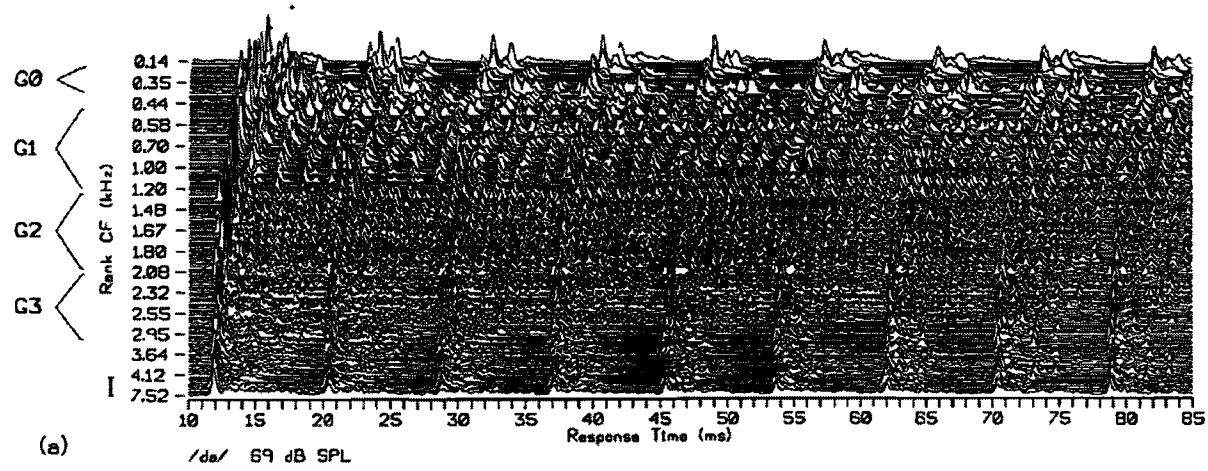

Figure 2-7: Firing rate vs. time of auditory nerve fibers in response to /da/ [reprinted with permission from (Secker-Walker and Campbell, 1990). Copyright 1990, Acoustical Society of America]

2.5 Pitch and Pitch Processing

In speech, the perception of pitch is dominated by the fundamental periodicity, i.e. by periodic repetitions at F_0 (Palmer and Shamma, 2004). There is still no consensus in the research community on how and where pitch is processed in the auditory system (Greenberg et al., 1987; Greenberg et al., 2004a; Palmer and Shamma, 2004). Studies using fMRI of the human auditory system have revealed that pitch is encoded at multiple stages of the ascending pathway (Griffiths et al., 2001; Penagos et al., 2004; Hall and Plack, 2008; Krishnan and Gandour, 2009).

There are two theories of pitch processing that are associated with the extraction of the fundamental frequency of complex stimuli: *rate-place* or *spectral pitch* and *temporal pitch* encoding. Rate-place neural models and the temporal models are used to explain pitch perception using either spatially defined neural maps or by time differences between neural discharges (Sachs and Young, 1979; Cariani and Delgutte, 1996). Under some conditions, each of these theories fails to explain observed pheonomena. For instance, the perception of temporal pitch is still present even when the energy at the fundamental is significantly reduced or when noise is added to the region surrounding F_0 (Cariani and Delgutte, 1996).

Modern theories of pitch processing incorporate combinations of the spectral and temporal neural mechanisms in the auditory system (Cariani and Delgutte, 1996).

There is also still uncertainty regarding the neural mechanisms involved in the encoding of pitch in situations where noise is present, especially when considering the enhancement of tuning characteristics of the neurons along the auditory pathways (both ascending and descending) as a result of a particular experience or exposure (Krishnan et al., 2010).

2.6 Auditory Evoked Potentials

Auditory evoked potentials refer to the aggregate neural electrical activity elicited in response to acoustic stimuli and recorded with electrodes placed on the scalp (figure 2-8). Measuring auditory evoked potentials is done to assess the integrity of the neural activation in the auditory system. Auditory evoked potentials that originate in the brainstem are referred to as Auditory Brainstem Responses (ABR), and can be measured using electrodes placed at specific locations. Generally, auditory evoked potentials are very low in amplitude (<1 μV) and are often buried under a variety of electrical signals originating from the subject or the surrounding environment, and which are substantially higher in amplitude. Auditory evoked potentials may be classified into two different types: transient, and sustained.

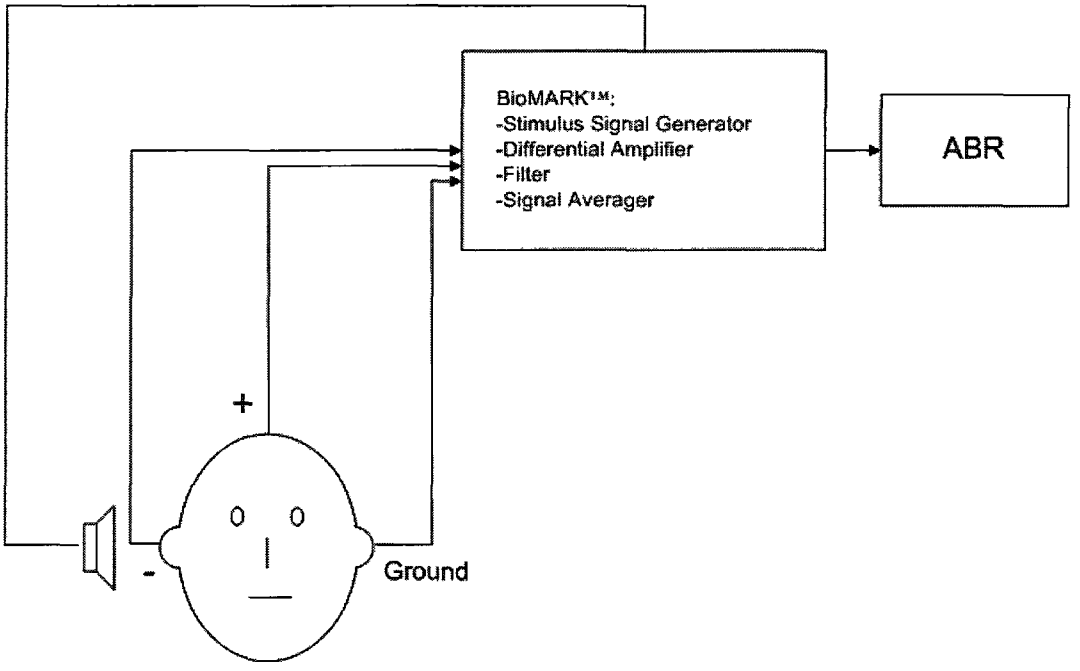

Figure 2-8: Diagram depicting the collection of auditory evoked potentials. When the electrodes are placed as shown, the main source of the potentials is the auditory brainstem, and the recorded signal is an Auditory Brainstem Response (ABR)

2.6.1 Auditory Evoked Potentials (Transient)

The auditory brainstem response includes the *transient response* generated 0-20 ms after the onset of the stimulus. Obtaining the transient response is a very common clinical procedure used in the evaluation of the auditory system, and is usually obtained using a tone burst or a click. The measurement of the interpeak latencies in the response gives an indication of how well the nuclei located along the ascending pathways respond to the acoustic stimulus. It also gives indication of how well the auditory system processes the acoustic stimulus when noise is present. For the transient response, specific relays are activated in a sequential fashion, located along the ascending auditory pathways. The transient response thus provides important information about the integrity of auditory signal processing from the auditory nerve up to the inferior colliculus.

When collecting the transient response using a click or tone burst as stimulus we obtain between five and six peaks that can be identified as follows (Sachs and Young, 1979; Hall, 2007c):

Peaks I and II: The first two noticeable peaks of the ABR recording come from the auditory nerve. The interpeak latency is usually 1 ms and they will occur within the first 2.5 ms after the onset of the stimulus.

Peaks III and IV: Studies report that they originate from the brainstem region. The interpeak latency is usually 1 ms and peaks III and IV will occur with a lag of 3-4 ms after the onset.

Peak V: This peak of generally low amplitude is generated by the inferior colliculus (ICC) with a lag of 6-10 ms after the onset.

Peak VI: This peak originates from the medial geniculate body of thalamus (MGB) and is generated after 10 ms from the onset.

Figure 2-9 shows an averaged ABR obtained using the speech stimulus /da/. The highlighted region indicates the location of the onset VA complex in the transient response.

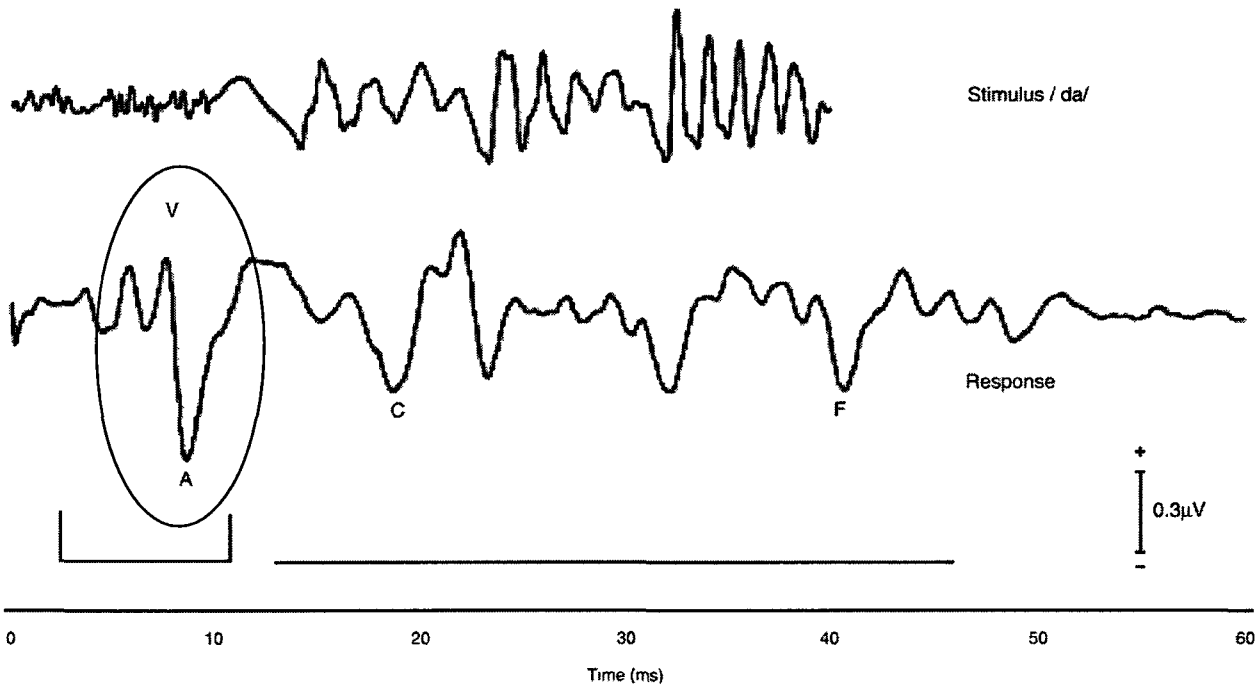

Figure 2-9: Depiction of the time-domain speech ABR obtained using the stimulus /da/ [reprinted with permission from (Russo et al., 2004). Copyright 2004, Elsevier.]

Note that the onset response to speech originates from the brainstem (response is found 5-10 ms after onset). The VA complex represents a transition between the peripheral system and the rostral (i.e. upper) brainstem (Chandrasekaran and Kraus, 2010). In reference to the Figure 2-9, wave C marks the beginning of the periodic, sustained portion of the response.

Previous studies using a five-formant synthesized speech syllable (/da/) have demonstrated that increasing the level of background noise had an effect on the latency of the onset responses referred to as V and A by increasing the delay and decreasing the slope of the VA complex. In response to noise, the latencies increase between the significant peaks of the transient response (Russo et al., 2004; Wible et al., 2004; Song et al., 2006).

2.6.2 Auditory Evoked Potentials (Sustained)

When using a periodic or steady acoustic signal as stimulus, an additional response to the transient response is obtained called the *Auditory Steady-State Response* (ASSR). The ASSR can be obtained using pure sinusoidal signals, frequency/amplitude modulated signals, or speech sounds and in particular vowels (natural or synthetic where the fundamental frequency and the formants are kept constant over time). The spectrum of an amplitude modulated sinusoidal stimulus signal contains the carrier frequency and the sidebands (F_c-F_m, and F_c+F_m). However, since the peripheral auditory system includes a non-linearity that can be approximated as a half-wave rectifier, it effectively demodulates the stimulus and a response is evoked at the modulation frequency (Geisler and Greenberg, 1986; Lins and Picton, 1995). In general, the sustained response obtained using a speech stimulus is similar but not identical to the standard ASSR. It is similar because the amplitude and frequency of the steady part of the signal are modulated at F_0 (Dajani et al., 2005), but it is also different because speech does not contain a single carrier frequency. In addition, if the speech stimulus is natural, then the amplitude and frequency of its spectral components are not perfectly steady over time (Picton et al., 2003). In the case of speech, the sustained response is referred to as either the Frequency Following Response (FFR) or the Envelope Following Response (EFR) (Aiken and Picton, 2006; Aiken and Picton, 2008). This response, which usually occurs starting around 20 ms after the onset of the stimulus, is believed to be a compound response originating from multiple generator sources located in the upper

brainstem region (Galbraith, 1994). However, there is uncertainty about the exact origins of the response (Chandrasekaran and Kraus, 2010).

2.7 Envelope Following Response (EFR) vs. Frequency Following Response (FFR) to a Speech Stimulus

An important distinction must be made between the Envelope Frequency Response (EFR) and the Frequency Following Response (FFR). The EFR is defined as the response to the envelope of the speech signal. The speech frequency components are modulated at the fundamental frequency (Dajani et al., 2005; Aiken and Picton, 2006). This concept explains why a response at F_0 is obtained when the fundamental frequency is missing from the stimulus. It is the entire frequency range of speech that contributes to the response at F_0 and not just the component F_0 in the stimulus. In practice, the EFR is obtained from the average between the response to the original stimulus and the response generated from the inverted polarity stimulus (i.e. (original+inverted responses)/2 denoted as '+-' configuration) (Darwin et al., 1989; Aiken and Picton, 2008). Inverting the polarity of a stimulus means that the order in which the eardrum receive compressions and rarefactions is reversed to produce different responses (Maurer et al., 1980).

Phase-locking is an important form of encoding inside the auditory system, because it provides essential information about the main components of speech sounds (fundamental frequency and formants) and provides robustness to noise. The envelope frequency response is generated from neural activity that is phase-locked to the envelope of the stimulus. As a result, the EFR occurs at F_0 and at some of its harmonics if the signal envelope is not sinusoidal.

To explain how the EFR at F_0 is generated, the peripheral auditory system can be seen to act as a half-wave rectifier where the envelope of the stimulus waveform is extracted. Since the fundamental frequency of the envelope is F_0, the response in the frequency domain shows components at F_0 and some of its harmonics. Figure 2-10 is a simplified version of the envelope extraction process at the peripheral level. In this example the carrier frequency is 1000 Hz modulated at 80 Hz with modulation depth of 100%. The stimulus passes through a

half-wave rectifier and on the right is the output representing the temporal firing pattern of an auditory nerve fiber. The response in the frequency domain shows a component at 80 Hz.

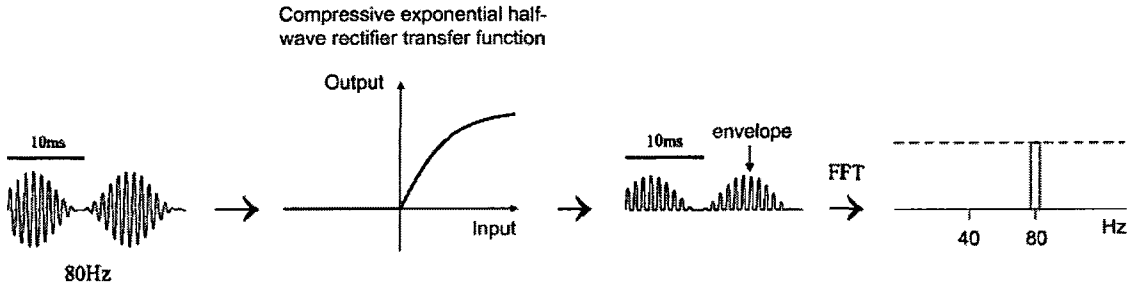

Figure 2-10: Simplified diagram of the extraction of the EFR in the cochlea [adapted from (Otavio and Picton, 1995) with permission from Copyright 2006, Elsevier.]

Figure 2-11 shows an example of an EFR obtained using the synthetic vowel /a/ (with F_0 at 100 Hz) as stimulus in quiet. The response at F_0 is clear along with its harmonics since the envelope of the stimulus is not purely sinusoidal.

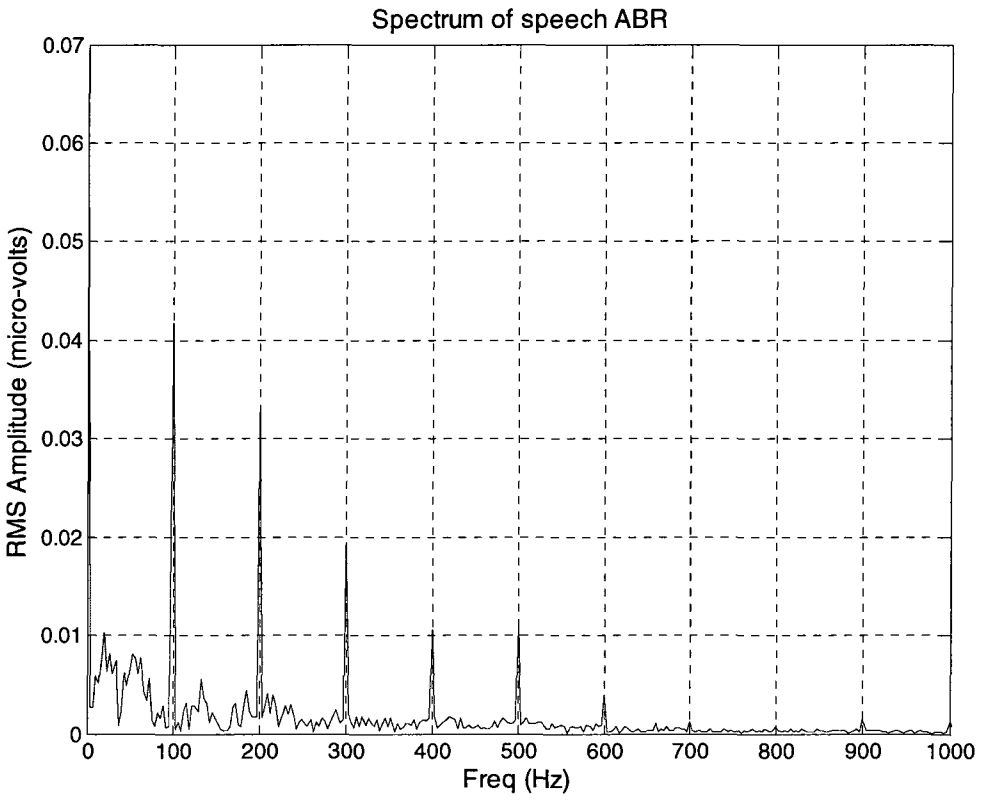

Figure 2-11: Example EFR to the /a/ vowel stimulus with F_0=100 Hz and F_1=700 Hz

On the other hand, the Frequency Following Response (FFR) follows the actual harmonics in the stimulus (and not the envelope). In practice, the FFR is obtained by subtracting the average response generated using the inverted polarity stimulus from the average response obtained using the original polarity stimulus: (original-inverted)/2 denoted as '--' configuration) (Aiken and Picton, 2008).

The steady-state Frequency Following Response (FFR) is primarily concerned with the responses around the formant frequencies, which in this study is restricted to the response in the region of F_1. The strength of the phase-locking activity from the auditory system to the harmonic components of the speech signal is reflected by the magnitude of the components of the FFR. This relationship might be due to the contribution of separate populations of auditory fibers having specific characteristic frequencies at or near the harmonic components of the stimulus (Sachs and Young, 1979; Krishnan, 2002). This place-specific activity in the cochlea would likely cause different populations of neurons at the brainstem level to synchronize to specific harmonic components through phase-locking (Bledscoe and

Moushegian, 1980; Krishnan, 2002). In the case of the FFR at F_1, a population of neurons with characteristic frequencies at or close to F_1 are likely the source of the response.

The response at F_1 is an FFR, but the response at F_0 is unlikely to be an FFR for three reasons: 1) When analyzing the data using the '--' configuration, we found that the response at F_0 is suppressed. 2) A study on EFR and FFR revealed the same observations (Aiken and Picton, 2006; Aiken and Picton, 2008). 3) When F_0 is missing from the stimulus, a strong response is still obtained at F_0 suggesting that the response at F_0 is evoked by the envelope of the stimulus (i.e. an EFR).

Figure 2-12 is an example of an FFR obtained using the vowel /a/ as stimulus in quiet where the response at F_0 is suppressed as expected, and the harmonics in the region of F_1 (700 Hz) are present.

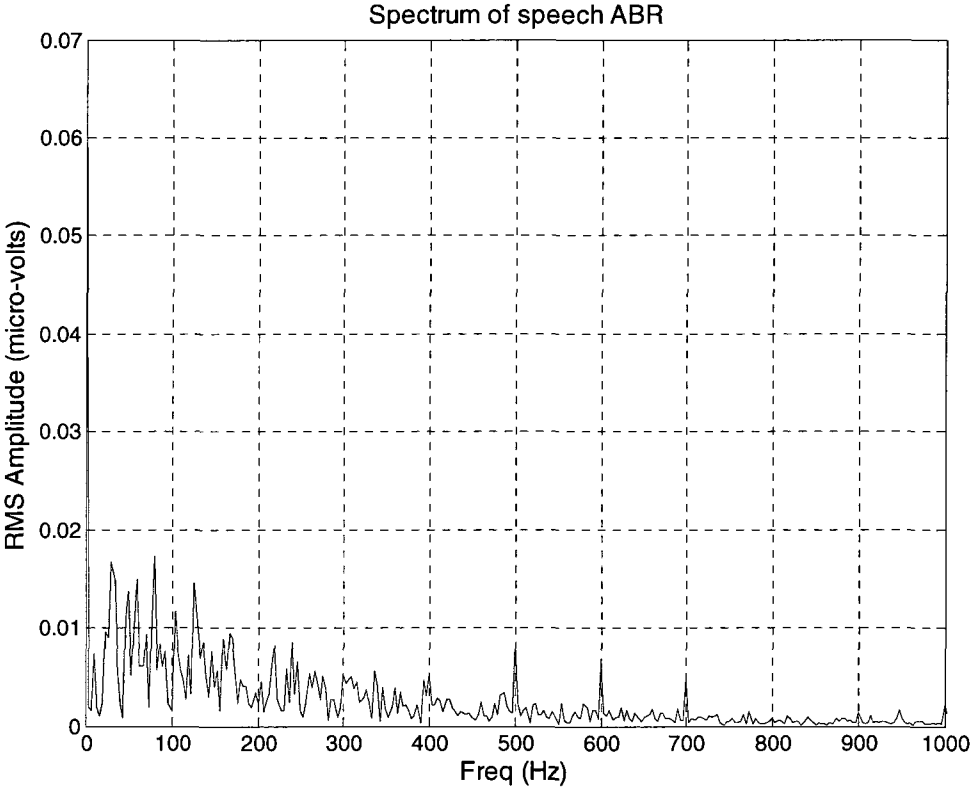

Figure 2-12: Example FFR to the /a/ vowel stimulus with F_0=100 Hz and F_1=700 Hz

2.7.1 EFR and FFR in Noise

When background noise is present in the stimulus, it decreases the contrast between the peaks of the formants and frequency troughs. It also adds random frequency peaks throughout the spectrum which could make weak formants undistinguishable from the surrounding peaks.

When stationary background noise such as multi-speaker babble is present over the speech stimulus, the spectral contrast (defined as the difference in dB between the peaks and their adjacent valleys) of the obtained speech ABR is reduced (especially for formants at higher frequencies). In addition, the shape of the spectral envelope varies making the overall response significantly different from the one when noise is not present in the stimulus (Assmann and Summerfield, 2004). The location of the formants identifying the vowel may be shifted due to the introduction of the noise (Cunningham et al., 2001; Assmann and Summerfield, 2004; Greenberg et al., 2004b; Palmer and Shamma, 2004; Russo et al., 2004; Wible et al., 2004; Johnson et al., 2005; Kraus and Nicol, 2005; Song et al., 2006). Russo et al. have found that adding background noise to a speech syllable like /da/ affected the filter components in the response (i.e. at the formants) more than that of the source (i.e. at F_0) (Russo et al., 2004). In the case of our designed stimulus, similar observations were expected.

At low SNR, the auditory system is often still able to encode the stimulus because of the ability of the hair cells to phase-lock to periodicities in the stimulus (Greenberg and Ainsworth, 2004). Phase-locking mechanisms allow the hair cells to phase-lock to the speech modulation frequency F_0 under various conditions, even when the fundamental frequency is missing. However, the processing of speech sounds embedded in noise may require various physiological processes relevant to neural encoding at various level of the brainstem, or large neural populations to encode the sound accurately (Kraus and Nicol, 2005; Song et al., 2006). It is expected that the EFR be more robust to noise than the FFR, because the EFR is generated from a broad range of frequencies modulated at F_0 as opposed to the FFR which only results from the energy at F_1 in the stimulus. Figures 2-13 and 2-14 show examples of FFR and EFR obtained using the vowel /a/ embedded in noise.

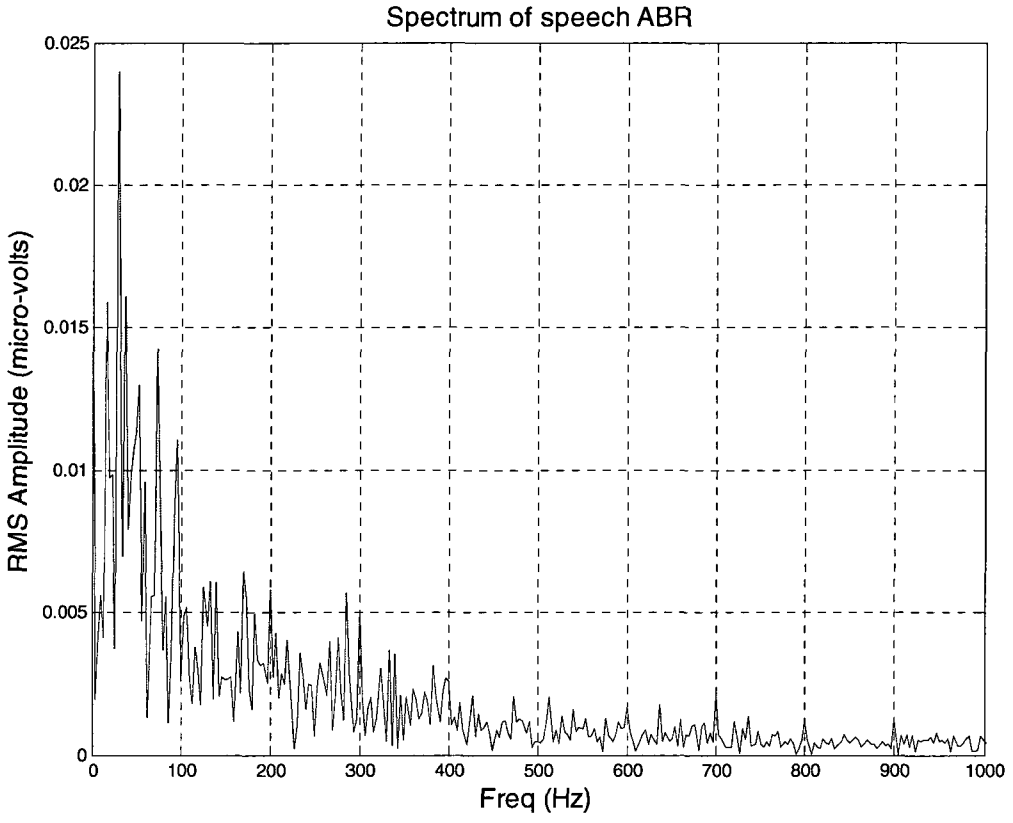

Figure 2-13: Example of FFR using vowel /a/ with SNR: -5dB

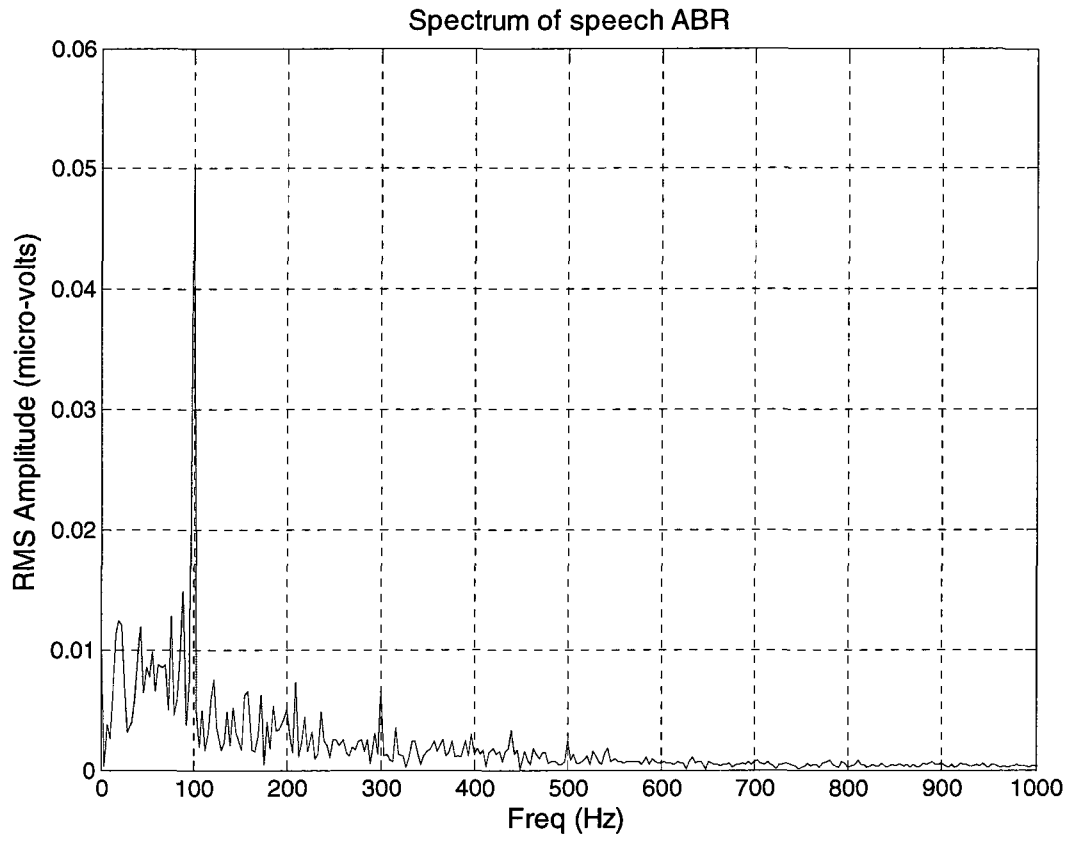

Figure 2-14: Example EFR using /a/ with SNR: -5dB

Figure 2-15 provides a graphical summary of the basis of the EFR at F_0 and the FFR at F_1, and how the frequency content (formants and resolved/unresolved harmonics) of the stimulus contribute to these responses. The first graph shows the spectrum of the stimulus shown in figure 2-3, where F_0 and the first three formants (F_1, F_2, and F_3) are identified. The FFR at F_1 is shown on the bottom left and the EFR at F_0 is on the bottom right; both are the grand-averaged responses obtained from this study. Arrows indicate which frequency components of the stimulus contribute to the response at F_0 and F_1. The FFR at F_1 depends on the component in the stimulus at F_1 while the entire frequency range (resolved and unresolved harmonics) contributes to the EFR at F_0.

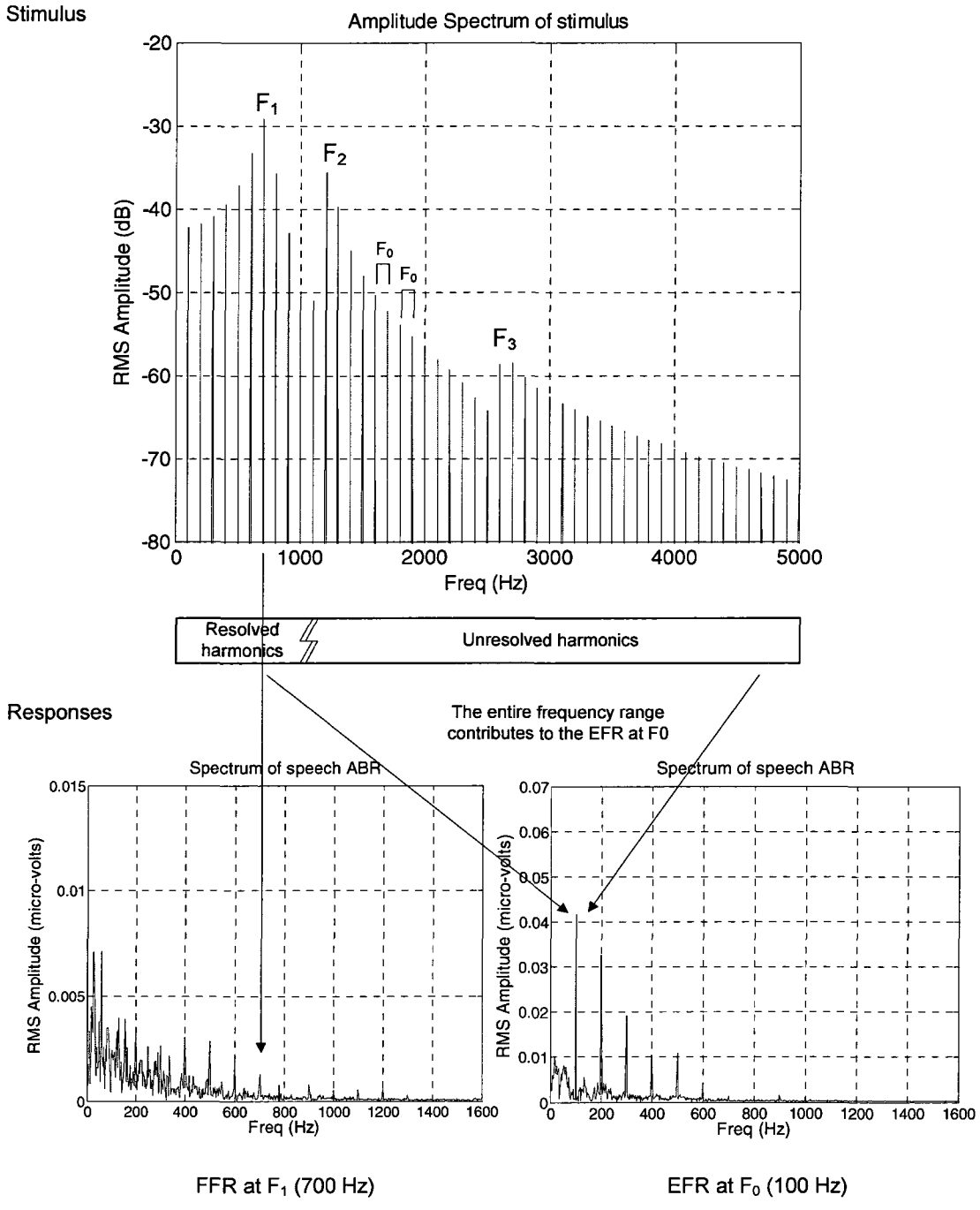

Figure 2-15: Block diagram describing the basis of the EFR at F_0 and FFR at F_1 (grand-averaged responses obtained from this study shown)

2.8 The Source Filter Model of Speech Processing

Kraus et al. (2005) have introduced a model describing auditory speech processing referred to as the *source-filter model of speech processing*. This model suggests that the source and the filter components of speech are processed along two separate pathways. Neural activity in these pathways gives responses that are associated with the sound generator (vocal cords) and the filter (oral cavity). The source is represented as the period between each of the major peaks of the speech ABR corresponding to the fundamental periodicity (i.e. the EFR); while the filter is represented by the transient peaks and the secondary peaks located in between each major peaks forming the FFR and corresponding to the first and possibly second formant (Kraus and Nicol, 2005). The source and filter evoked responses are used to evaluate the integrity of the central auditory system. For example, they can be used in clinical assessment of the auditory system of children having learning disabilities and in their rehabilitation, as there is evidence that they are independent and degrade differently with learning impairment (Kraus and Nicol, 2005). Improvements of the filter response were found in children having language-based learning problems who followed auditory training programs. In addition, increasing the level of background noise in the speech stimulus significantly deteriorated the filter response while the neural representation of the source was little changed. These two lines of evidence support the presence of two pathways for processing the source and filter, at least up to the upper brainstem.

2.9 Cortical Speech Evoked Potentials

It is possible to assess the neural activity in the auditory cortex in response to a speech stimulus using cortical speech-evoked potentials (Martin et al., 2008). One of the cortical potentials clinicians record for this purpose is Mismatch Negativity (MMN). In contrast to the ABR where the same stimulus is presented repeatedly, MMN is based on acoustic differences in the stimulus among a series of homogenous stimulus presentations (Kraus et al., 1993; Kraus et al., 1995b). The difference in the stimulus sequence (i.e. intensity, duration, frequency) produces cortical-related evoked potentials that can be measured, and provide insight into the efficiency of neural encoding of speech at the cortical level. Hence, MMN does not provide information about the processing of the content of the speech

stimulus itself as do brainstem speech evoked responses, but gives an indication as to how the auditory cortex processes acoustic contrasts which relate to speech discrimination and speech perception. Cortical responses compared to brainstem responses are much lower in frequency (<10 Hz) and significant events are separated by several milliseconds (Cunningham et al., 2001). Moreover, the auditory evoked potentials measured at the cortical level are of greater amplitudes than the ones from the brainstem, because the auditory cortex is a much larger structure than any one making up the brainstem, and is closer to the recording electrodes (Burkard et al., 2007).

2.10 Methods of Speech Synthesis

Artificial speech synthesis is commonly found in Text-To-Speech systems (TTS) such as automated telephone systems or as support for people with speech disorders. There are two main methods used to produce synthesized speech: *Concatenative* and *Formant* synthesis. In concatenative synthesis, messages are formed using pre-recorded human speech segments by 'concatenating' them together (Van Santen et al., 1997; Holmes and Holmes, 2001). The digitally pre-recorded segments could be words or diphones stored in a database. Diphones contains two consecutive phonemes, which are the smallest linguistically distinctive unit of speech (Dutoit, 1997; Van Santen et al., 1997; Holmes and Holmes, 2001). Formant synthesis is the second most popular method used to synthesize speech, and is the one used in this study to produce the stimulus vowel /a/, and its variants. Formant synthesis is based on the *source-filter model of speech production.* Contrary to the concatenated technique of speech synthesis, formant synthesis is not restricted to one voice. It has the flexibility of producing female, male, old and young voices (Holmes and Holmes, 2001; Eduardo, 2002). However, this technique involves a certain level of complexity and requires specific parameters related to amplitude, frequency and filtering. Dennis H. Klatt was a pioneer in speech and hearing science and produced the first generation of formant synthesizer in 1980. Still used today in research on speech, the Klatt synthesizer produces relatively human-like speech sounds (vowels and consonants) using a particular set of parameters as input: formants, bandwidths of resonators, and pitch frequency. For synthesizing the vowel /a/ and its variants in this thesis, a simplified form of Klatt's formant synthesizer restricted to generating steady-state vowels was implemented.

Chapter 3 : Methodology

This chapter presents the general framework of the research methodology. A detailed description of the participating subjects, and the methods employed in generating and collecting the speech ABRs are included. The experiment protocol and the steps involved in the construction of the stimuli follow, and details about the calibration and the tests performed to evaluate the quality of the recorded signals are also presented here. A brief description of coherent averaging and its importance in the measurement of ABRs is included as well. Finally, the chapter concludes with a description of the statistical analysis and set-up of the ANOVA.

3.1 Subjects

Twelve normal hearing subjects took part in this study (six females and six males). The subjects ranged in age between 25 and 40 yrs old. At the start of each session, each participant was asked to read and sign an informed consent form. Any questions that the participants had were answered and verification was made with each participant to make sure they understood the content of the consent form. This study was carried out in accordance with the Research Ethics regulations that are overseen by the University of Ottawa Research Ethics Board.

Participants were asked to sit on a comfortable reclining chair located in an acoustical booth at the University of Ottawa Health Campus, in the School of Rehabilitation Sciences (Audiology Program). Each subject was asked to minimize movement during the recordings. To keep them awake but not tense, the subjects watched a muted movie with subtitles on a computer screen installed in the booth. Each participant underwent an audiometric test prior to the experiment to verify if their hearing threshold was below 15 dB HL (using pure tone air-conduction test) at 500, 1000, 2000, and 4000 Hz. The audiometric test was performed using an Interacoustics® AC40 Clinical Audiometer.

3.2 Experiment Protocol

This study involved the presentation of three variants of the synthesized vowel /a/ as in 'hod' in quiet and in noise. These three variants consisted of the first three formants (Allformants), the first formant only (F_1Only), and the second and third formants only (F_2&F_3Only). In the noisy stimuli, stationary white Gaussian noise was added (at an SNR: -5dB), so in total, there were responses to six stimuli obtained from each participant.

The first objective was to study the responses using the clean and noisy stimuli. Statistical analysis was done on the collected steady-state speech evoked auditory brainstem responses at the fundamental frequency F_0 (100 Hz) and at the first formant F_1 (700 Hz). A quantitative and qualitative analysis was performed on the transient response as well.

The second objective aimed at investigating the contribution from the frequency regions of the vowel dominated by resolved and unresolved harmonics to the response at F_0. This was done by presenting a stimulus consisting predominantly of resolved harmonics (F_1Only), a stimulus consisting predominantly of unresolved harmonics (F_2&F_3Only), and the stimulus containing the first three formants together (Allformants). The order of presentation of the stimuli was randomized for each participant. Statistical analysis was done to compare the steady-state responses at F_0 to these three variants, in quiet and in noise.

In summary, the following stimulus variants were presented to each subject:

Quiet	Noisy
Allformants	Allformants
F_1Only	F_1Only
F_2&F_3Only	F_2&F_3Only

Table 3-1: Simuli used for each subject

In this experiment, the response at F_2 was not analyzed because it was beyond the bandpass filter limits that were set to 30-1000 Hz on the BioMARK™ system used to generate stimuli and record responses. The amplitude of the response at F_0 obtained from the Allformants stimulus was expected to be greater than the ones obtained from the F_1Only or F_2&F_3Only stimuli regardless if the stimulus was noisy or not, because in the former case both the

resolved and unresolved harmonic regions contribute to the EFR. The responses at F_1 from the Allformants and F_1Only stimuli were expected to be similar because both stimuli have the same content at F_1, and consequently should have similar FFRs. The FFR at F_1 from the F_2&F_3Only variant was expected to be weaker given that the energy in this stimulus at F_1 is weak.

3.3 Creation of the Stimulus and Recording of the Response

The speech auditory evoked brainstem responses were collected using the BioMARK™ v.7.0.2 system (Biological Marker of Auditory Processing, Bio-logic Systems Corporation). This system, which is used as a clinical tool for measuring auditory evoked responses, was originally developed by Nina Kraus and her colleagues at Northwestern University (IL, USA), but is now manufactured by Natus Corporation. The ABR recording system contains a USB hardware interface (Bio-logic® Navigator® Pro), which sends the acoustic signal through insert earphones (Bio-logic®) and collects the evoked-responses using three electrodes placed on the scalp of a subject. For this experiment, the stimulus was delivered to the right ear through a Bio-logic® small foam insert earphone. The left ear was kept un-occluded for the entire experiment. The montage of the active electrodes consisted of a single channel with two inputs (a signal and a reference), and a grounded electrode. To record a strong and robust speech ABR, a vertical montage (vertex (Cz) to ipsilateral earlobe) was used as opposed to a horizontal montage (earlobe to earlobe) (Chandrasekaran and Kraus, 2010). The montage was also the one suggested by the BioMARK™ system when using a single channel with speech stimulus. The first (signal) recording input was located at Cz (vertex) using a non-disposable Ag-AgCl scalp electrode. The second (reference) recording input was located on the right earlobe using a disposable electrode. The ground electrode was placed on the left earlobe.

The skin of the participants was prepared using mild abrasive (Nuprep from D.O. Weaver®) and with Ten 20 conductive EEG paste (D.O. Weaver®) which was used only for the non-disposable electrode, in order to achieve single electrode impedance of 5 kOhms with inter-electrodes impedances of less than 1 kOhm. The impedances were recorded in a logbook and they were verified before each recording and immediately after. If the single electrode

impedance was above 10 kOhms then the recording was discarded. For each stimulus, the obtained response was band-passed on-line to suppress EEG noise. The band-pass filter had cutoff frequencies of 30 and 1000Hz to allow recording of the responses at F_0 (100 Hz) and at F_1 (700 Hz). The recordings were obtained with 1024 sampling points and with a recording epoch time as close as possible to the duration of the specified stimulus. For the stimulus of 300 ms duration, the epoch time was 319.8 ms giving a sampling frequency of 3202 Hz (1024sampling points/319.8ms).

The recordings were obtained using stimuli with alternating polarity presented at 48 kHz with a 16-bit resolution. With alternating polarity, one stimulus presentation is produced by rarefaction of the acoustic tranducer's membrane, and the next with condensation (Akhoun et al., 2008b). Averaging the responses to the two polarities minimizes response artifacts and the cochlear microphonic (CM) signal (Johnson et al., 2005), which is an electrical signal generated by the cochlea's hair cells in response to an acoustic stimulus, and which occurs immediately before the ABR (Chandrasekaran and Kraus, 2010). As discussed earlier, this approach is used to estimate the EFR at F_0, while subtracting the responses to the two polarities allows the estimation of the FFR at F_1 (Aiken and Picton, 2008). Artifacts due to eye-blinks or other motion artifacts were eliminated on-line by discarding epochs in which the response exceeded 23.8 μV. The recordings were synchronized to the stimulus onset and delivered at a rate of three stimuli per second. With each stimulus condition, 3000 repetitions of the stimulus were presented over which the responses were coherently averaged.

In order to visualize and analyze the frequency content of the speech ABRs using Matlab® version 7.5.0 (The MathWorks, Inc), the raw data obtained from the system was converted to ASCII format using a software application provided with the equipment.

3.4 Stimulus

A 300 ms male formant synthesized speech stimulus (vowel /a/ from the English language) with a fundamental frequency (F_0) of 100 Hz was generated. The formant parameters were taken from the literature (Peterson and Barney, 1952). The first three formants (F_1 to F_3) were specified at 700 Hz, 1220 Hz, and 2600 Hz respectively (Klatt, 1980). The bandwidths

for the first three formants were specified as 130 Hz, 70 Hz, and 160 Hz respectively, and the relative amplitudes of F_2 and F_3 were specified at –4 dB and –27dB relative to F_1. This vowel was chosen as the stimulus in part because F_0 and F_1 are relatively well separated in frequency. Since the envelope at F_0 is not purely sinusoidal, and this study distinguishes between the FFR at F_1 and the EFR at F_0 and its harmonics, the separation in frequency between F_0 and F_1 in the stimulus diminishes the potential for interaction between the two responses. In particular, the response at F_1 (700 Hz) is expected to be dominated by the FFR, with minimal contribution from the EFR at the 7^{th} harmonic of F_0 (Aiken and Picton, 2008). Having the F_1 frequency as a harmonic of F_0 is not relevant when studying the FFR since this response follows the harmonic component (of F_0) closest to the formant frequency. For example, if F_1 is at 720 Hz we would expect a response at the frequency closest to the harmonic of F_0 (i.e. 700 Hz).

This vowel is also suitable to investigate the relative contribution of the frequency regions of the stimulus dominated by resolved and unresolved harmonic by utilizing two variants the first of which contains F_1 only and the second contains F_2 and F_3 only. In addition, this vowel allows comparison with previous studies where a single vowel was used as a speech stimulus, alone or as part of a consonant-vowel syllable (Dajani et al., 2005; Skoe and Kraus, 2010).

As mentioned above, both quiet and noisy stimuli were used. The quiet stimulus was generated by specifying a very high SNR (of 10000 dB) in the synthesizer program, and the noisy stimulus corresponded to an SNR of -5dB for the Allformants variant. The latter was preferred over a noisier SNR (e.g. -10 dB), because it eliminated the need to significantly increase the number of averages (realizations) needed to allow the detection of the response.

The stimuli were created using a synthesizer based on a simplified version of the Klatt cascade/parallel formant synthesizer (1980) (see figure 3-1), and were saved in wav format at 16-bit resolution (mono) and sampled at 48 kHz. A train of impulses at the fundamental frequency represented the source. For each formant, a resonator was designed acting as a narrow band-pass filter having a center frequency equal to the formant frequency for which it

was designed, and bandwidth as specified above. The train of impulses was passed through each resonator independently and the outputs were summed together afterwards.

Each digital resonators had a transfer function defined by the Klatt formant synthesizer (Klatt, 1980):

$$T(f) = \frac{A}{1 - Bz^{-1} - Cz^{-2}}$$

$$\text{with } z = \exp\left(\frac{j \times 2 \times \pi \times (\text{formant freq})}{Fs}\right)$$

where the constants were:

$$C = -\exp\left(\frac{-2 \times \pi \times BW}{Fs}\right)$$

$$B = 2 \times \exp\left(\frac{-\pi \times BW}{Fs}\right) \times \cos\left(\frac{2 \times \pi \times (\text{formant freq})}{Fs}\right)$$

$$A = 1 - C - B$$

with BW being the formant bandwidth specific to F_1, F_2 and F_3 given by (Klatt, 1980).

From this point, different signals were generated with the components that were needed (one containing all three formants, one with the first formant only, and one with the second and third formants only). For the stimuli containing all three formants or the second and third formants only, the outputs of the resonators were combined linearly. Note that the frequency and amplitude of formants may have shifted slightly from the specified values when combining the outputs of the resonators because of the overlap in their frequency responses. The choice of the duration of the stimuli (300ms) results in an integer number of periods at the fundamental frequency and its harmonics, which allows for precise Fourier analysis in the frequency domain because there is no leakage to adjacent bins. This duration was also chosen because of similarity to natural vowels, and the domination of the steady-state response versus the transient part in the ABR.

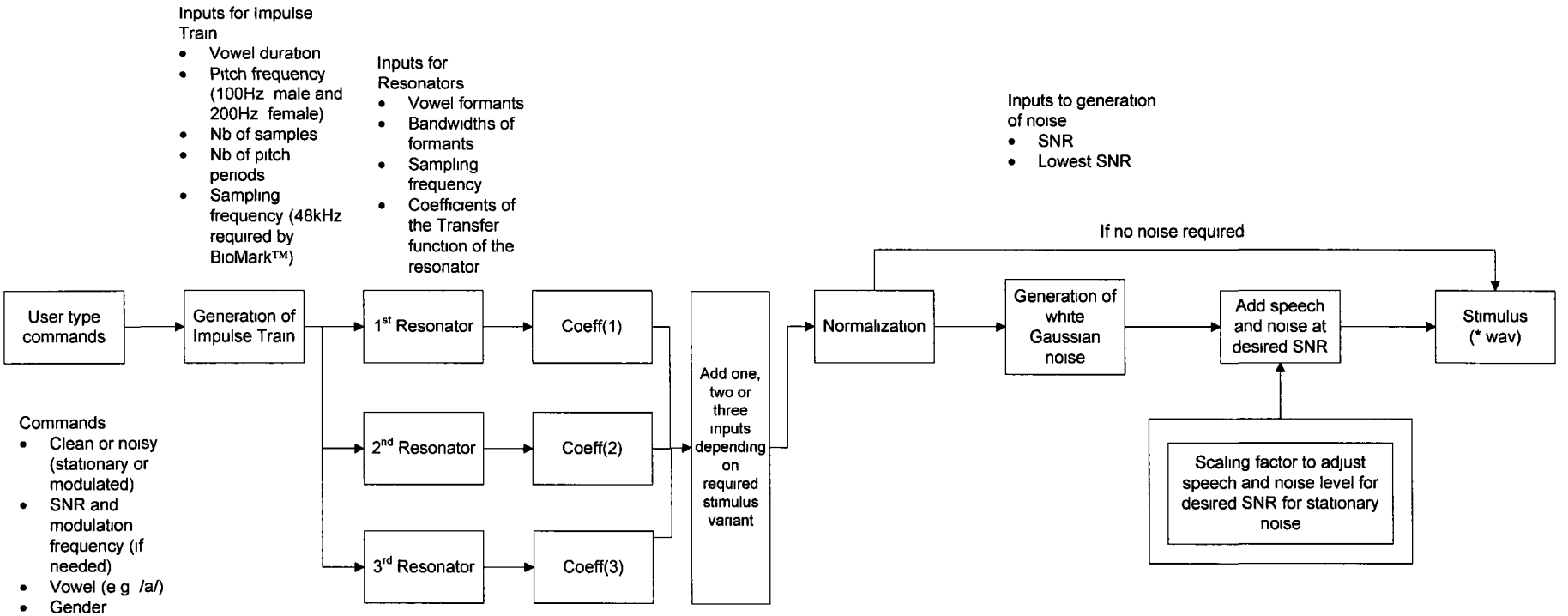

Figure 3- 1: Block diagram of synthesizer used to generate the stimulus signal

For the noisy stimuli, white Gaussian noise was generated only once and it was added to the speech signal at the specified SNR level for the Allformants variant. The resulting SNR was verified to match the specified SNR level. Once the noise level was adjusted to meet the specified SNR level, the same noise vector was used in all the three noisy conditions. Using a pre-determined noise vector for the entire experiment eliminates a potential source of variability by ensuring a consistent effect of the noise on the evoked responses regardless of the underlying speech stimulus. To eliminate the possibility that predictability of the noise has an effect on the statistical group comparisons between the responses under the different conditions, the order of the presentation of the stimuli was randomized to make it different for each subject.

The stimulus SNR was defined as follows:

$$SNR = 20 \times \log_{10}\left[\frac{\text{power of signal over entire duration}}{\text{power of noise over entire duration}}\right]$$

The synthesizer was designed so that the underlying speech stimulus had the same amplitude for both the quiet and noisy condition. Therefore, adding the noise increased the level of the stimulus. Normalization was performed to make sure the overall signal in the case of noisy condition, adjusted to the lowest SNR, covers close to the maximum range allowed by the .wav format of -1 to 1.

For each stimulus generated, the vowel spectrum clearly showed the harmonics of the fundamental frequency and the three formants (such as in figure 2-3). After verifying the relative amplitudes of the formants, each .wav file was used as a stimulus using the BioMARK^{TM} system.

Since the noisy stimulus is higher in level than the quiet stimulus, calibration was performed in reference to it, to control the sound level of the output and to keep it within a safe range (Appendix A). The stimulus intensity levels were calibrated with the earphone connected to a 2cc coupler attached to a Brüel & Kjaer Artificial Ear type 4152, and a Sound Level Meter (SLM) Type 2230. The calibration was set so that the signal level heard by the subjects was less than or equal to 80.0 dB SPL as measured using the SLM. To achieve this intensity

level, we adjusted an internal calibration factor in the BioMARK^{TM} system as needed. Once the desired level was obtained, it remained unchanged for the rest of the experiment.

Verification of the signal at the output of the right-ear earphone was done to check if there was any substantial distortion introduced by the system (BioMARK^{TM} + cables + earphone). This involved recording the signal obtained at the output of the SLM with the earphone connected to the 2cc coupler attached to the SLM as described above. The signal was digitized using a soundcard at a 44100 kHz sampling frequency, and 16-bit resolution. The recorded signal was then plotted in the time and frequency domains. This extra step confirmed that the shape of the original signal in time domain was not altered significantly by the transfer function of the system, and that the relative amplitudes of the formants obtained from the frequency spectrum were not affected significantly.

3.4.1 Time and Spectral Domain Analysis of the Stimuli Presented to the Subjects

These signals were recorded using one insert earphone (Bio-logic) inserted into a 2cc coupler connected to a sound level meter (SLM). A connecting cable was used to link the SLM microphone output to a soundcard. The signals were digitally recorded at 44100 Hz with 16-bit resolution and are shown in figures 3-2 to 3.13.

Quiet
Allformants:

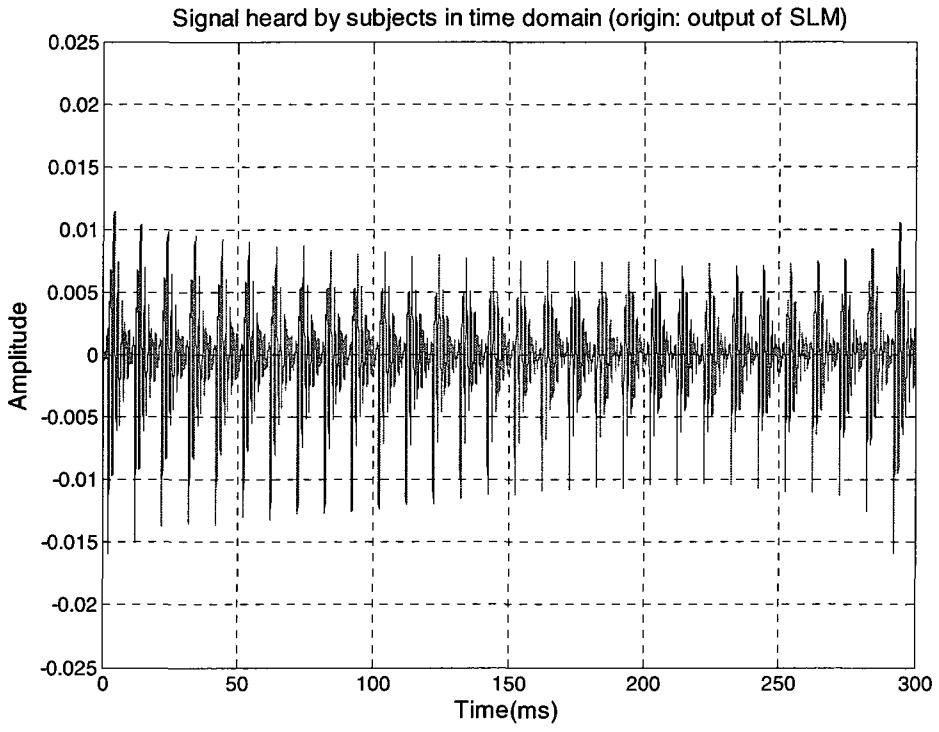

Figure 3- 2: Time domain output of SLM (quiet, Allformants)

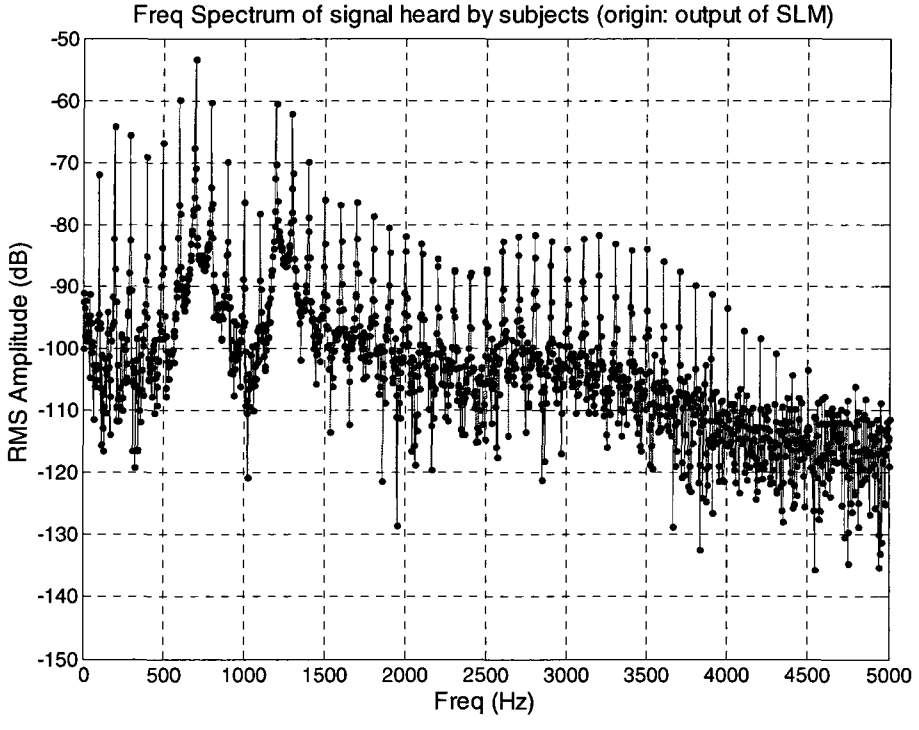

Figure 3- 3: Frequency spectrum of output of SLM (quiet, Allformants)

F_1Only:

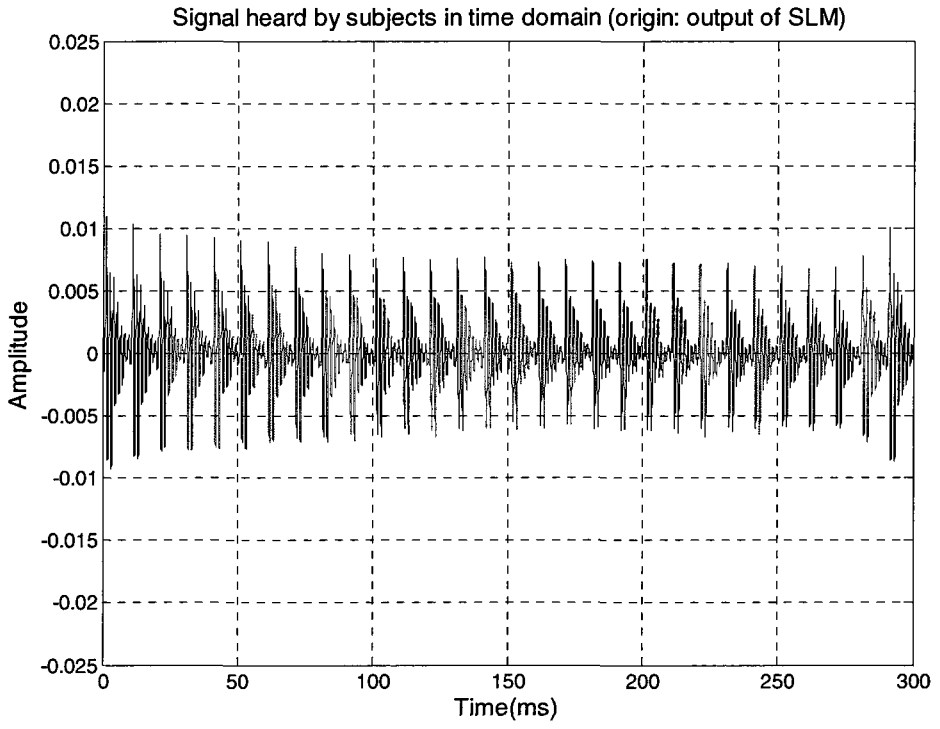

Figure 3- 4: Time domain output of SLM (quiet, F_1Only)

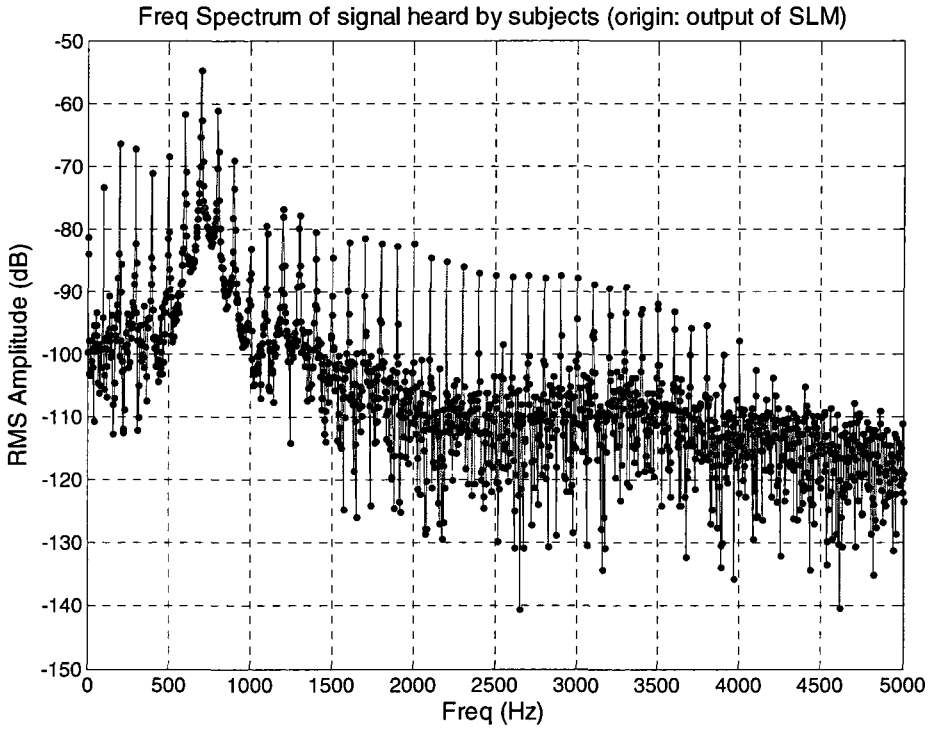

Figure 3- 5: Frequency spectrum of output of SLM (quiet, F_1Only)

F_2&F_3Only:

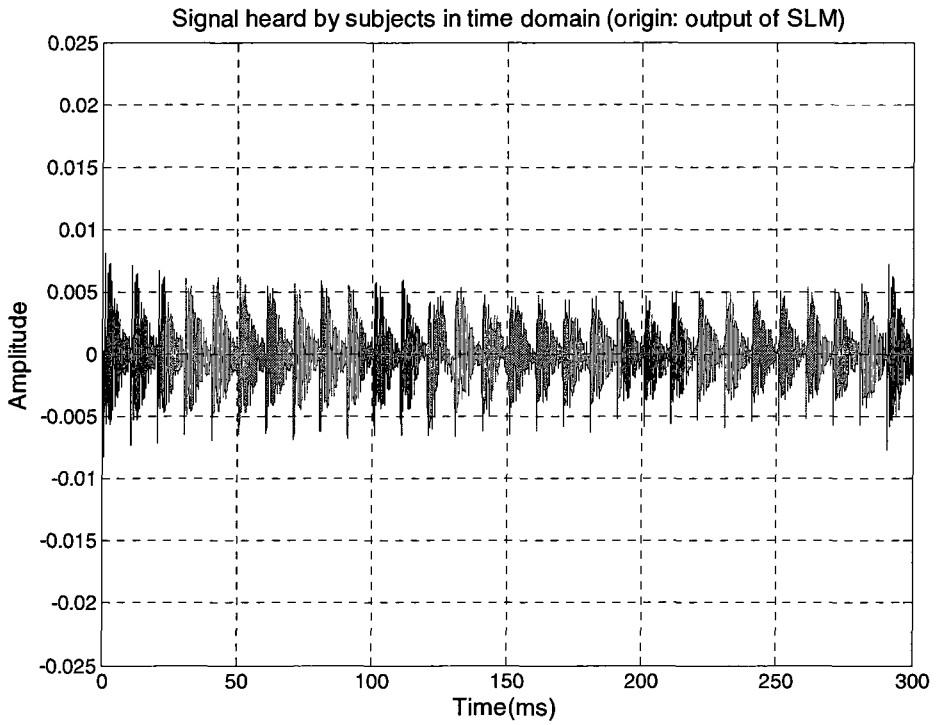

Figure 3- 6: Time domain output of SLM (quiet, F_2&F_3Only)

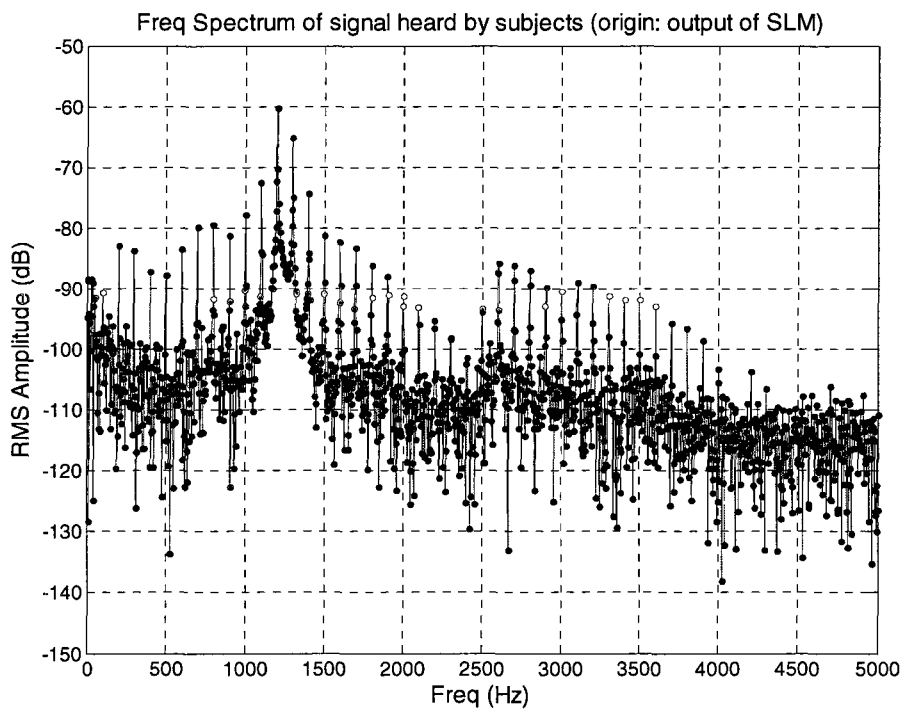

Figure 3- 7: Frequency spectrum of output of SLM (quiet, F_2&F_3Only)

Noisy
Allformants:

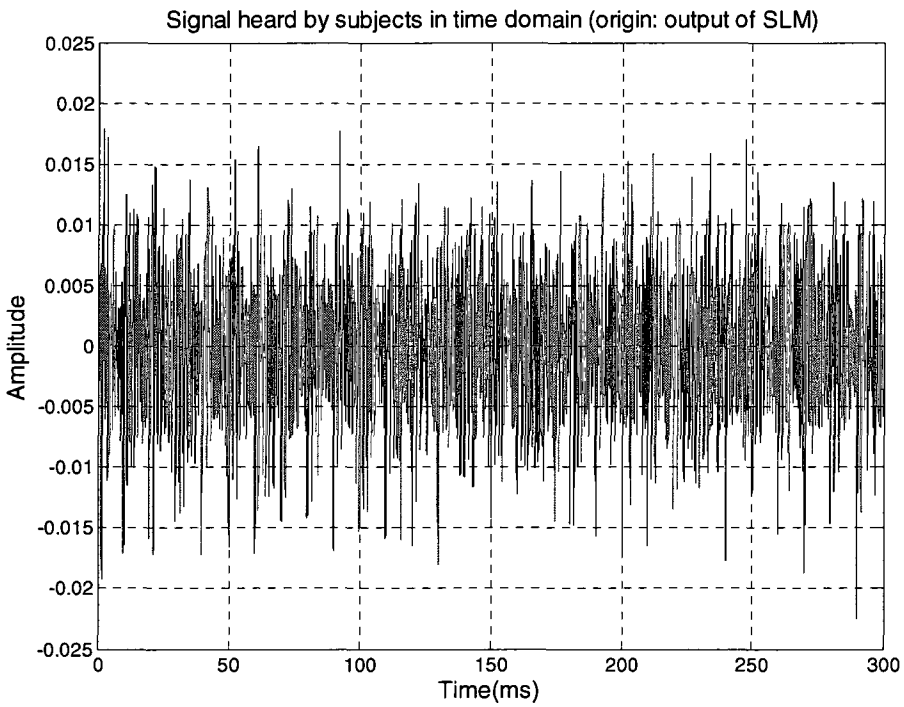

Figure 3- 8: Time domain output of SLM (noisy, Allformants)

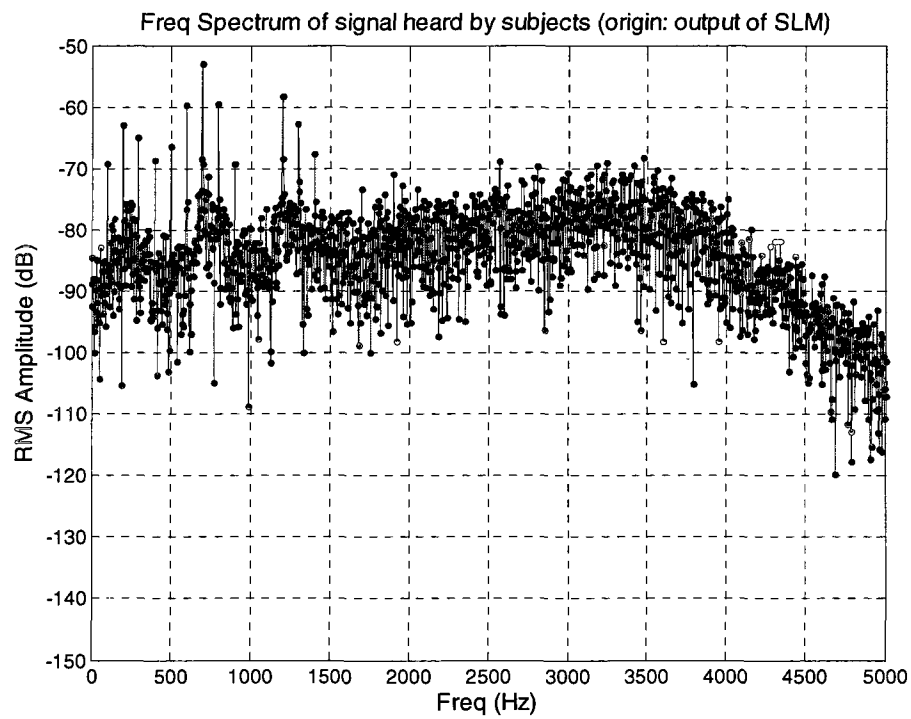

Figure 3- 9: Frequency spectrum of output of SLM (noisy, Allformants)

F_1Only:

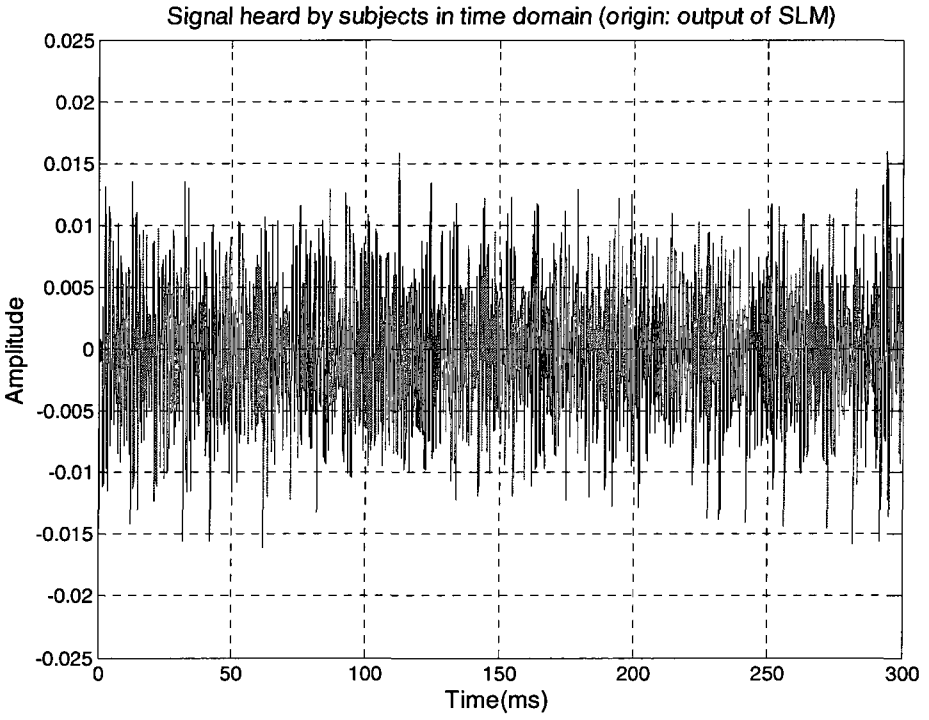

Figure 3- 10: Time domain output of SLM (noisy, F_1Only)

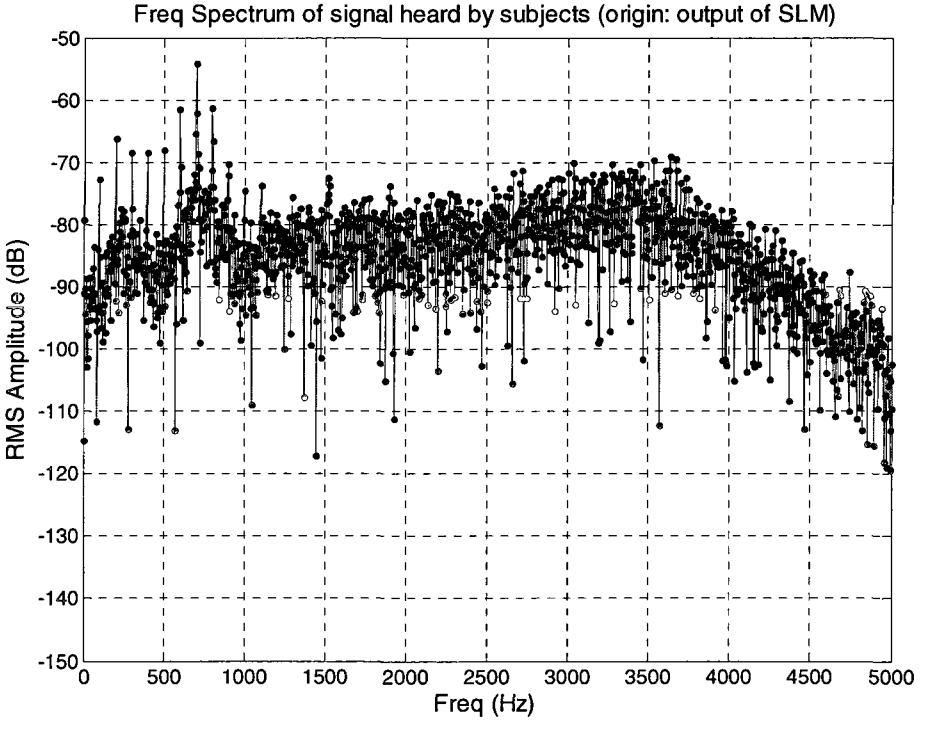

Figure 3- 11: Frequency spectrum of output of SLM (noisy, F_1Only)

F_2&F_3Only:

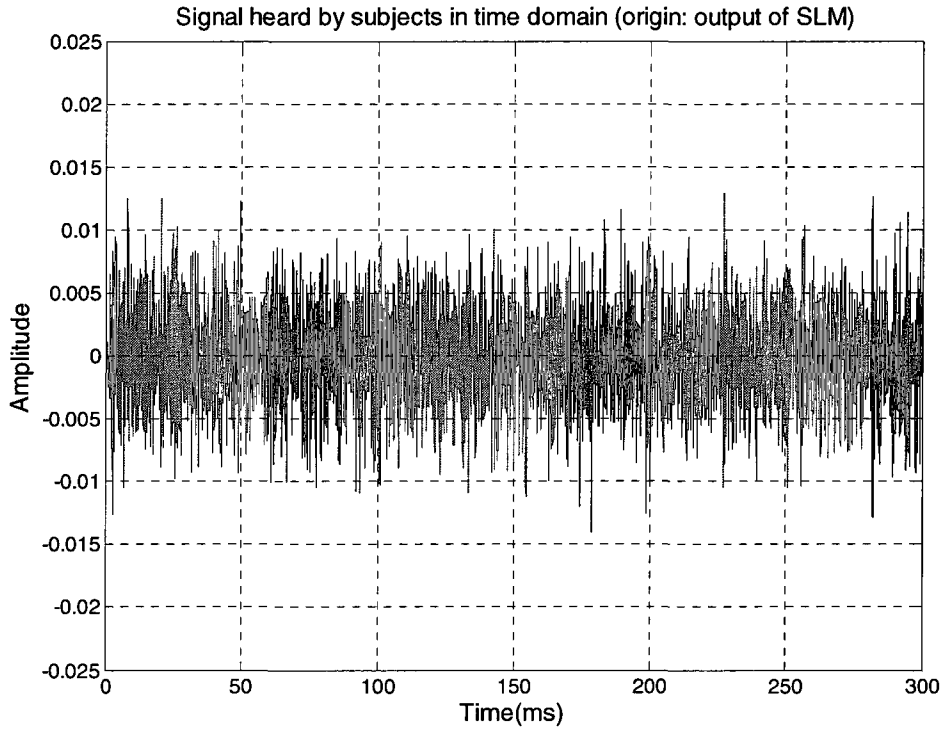

Figure 3- 12: Time domain output of SLM (noisy, F_2&F_3Only)

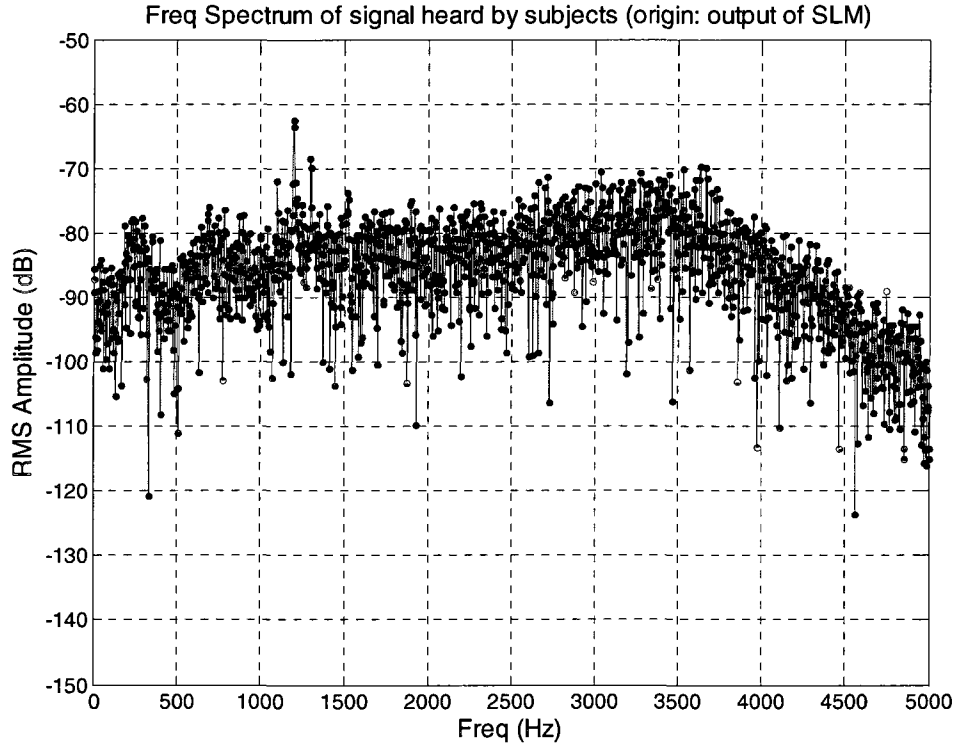

Figure 3- 13: Frequency spectrum of output of SLM (noisy, F_2&F_3Only)

3.5 Coherent Averaging

The concept behind coherent averaging is to reduce the noise variance by averaging recordings synchronized to the stimulus onset over multiple realizations. The higher the number of realizations of the stimulus, the higher the SNR of the averaged ABR becomes (Deutsch and Micheli-Tzanakou, 1987). This is an important step before starting the statistical analysis because the auditory evoked response is buried within other unrelated brain activity and other electrical signals originating from the subject or from the surrounding environment (Hall, 2007d). It is worth mentioning that cortical responses require less averaging (75-100) than the brainstem response (>1000), because they are generally larger in amplitudes (Chandrasekaran and Kraus, 2010). The EEG noise that represents brain activity unrelated to the stimulus is usually very high, resulting in a very low SNR in each recorded response. Therefore, using a noise reduction technique such as coherent averaging is a must.

The time domain coherent average is expressed as:

$$\text{coherent average} = \frac{\sum_{i=1}^{m} x_i(t)}{m}$$

where $x_i(t)$ represents the time domain response signal in a given realization, and m is the number of realizations (3000 in this case).

3.6 Signal Analysis

The analysis of the responses to the synthesized speech vowel consisted of five parts and focused on qualitative and quantitative aspects of the responses. The first aspect of the analysis focused on the averaged transient responses across all the subjects for the Allformants stimuli (clean and noisy). The V and A onset peaks result from transient events and not from the periodic acoustic content of the stimulus (Kraus and Nicol, 2005). They have been studied because they are the most robust positive and negative peaks of the transient response, and have been shown to be reliable and informative (Johnson et al., 2005; Kraus and Nicol, 2005). A comparison of the transient peak amplitudes and slope of the VA onset complex for both noisy and clean stimuli was performed.

Measuring the slope of the VA onset gives an indication of how much the responses degraded due to noise, as studies have shown that the addition of noise to the stimulus results in a shallower VA onset slope (Wible et al., 2004). An increased time delay of the VA onset for the noisy condition was also expected, as seen in the literature (Johnson et al., 2005; Kraus and Nicol, 2005). The estimation of the slope of VA was done manually by an expert in clinical auditory electrophysiology, on the grand-averaged response of all the subjects, since the responses from individual subjects were too noisy for visual analysis.

The second part of the analysis focused on the amplitude of the steady-state response at F_0 and at F_1 produced by the three stimulus variants, for both the noisy and quiet conditions. The response at F_0 is likely mainly an envelope following response (EFR), while the response at F_1 is likely mainly a frequency following response (FFR). Hence we used the so-called '+-' configuration, in which response to the two presentation polarities are averaged, to analyze the response at F_0, and the so-called '--' configuration, in which the difference between the responses at the two polarities is averaged, to analyze the response at F_1. The justification for using these configurations for detecting each of the EFR and FFR is discussed in detail in (Aiken and Picton, 2008). Estimation of the amplitude at F_0 and at F_1 was done over the entire steady-state response starting at 20 ms after the onset of the response, and so a Discrete Fourier transform was performed on the sustained portion of the response in the final 300 ms. With a bin width of approximately 3.33 Hz, the response components at F_0 (100 Hz) and F_1 (700 Hz) fall exactly into a single bin, with no leakage of energy into the surrounding bins. For each condition (three stimulus variants in quiet and in noise) the mean subject amplitude, the standard deviation, and the standard error of the mean were also computed.

Below is a table reporting the responses that were expected from each stimulus.

	Expected responses	
	F_0	F_1
Allformants	Yes	Yes
F1Only	Yes	Yes
F2&F3Only	Yes	No

Table 3-2: Response expected from stimulus variants

For all stimulus variants, an EFR at F_0 is expected, because all the stimuli have a fundamental periodicity at F_0. On the other hand, a FFR at F_1 is expected for the stimulus variants AllFormants and F_1Only. There is no (or little) FFR expected at F_1 for the F_2&F_3Only stimulus, because there is little energy at F_1 in this stimulus.

The third part of the analysis focused on studying the local rms-SNR of the responses at F_0 and F_1 produced by the three stimulus variants, for both noisy and quiet conditions. This was done by measuring the rms level at F_0 and F_1 relative to the noise bins surrounding it. A linear rms estimate of the SNRs is used (instead of a logarithmic one) in order to allow statistical comparisons of the means. A similar approach for evaluating the local SNR (mean noise-window SNR: nwSNR) was introduced by (Zhang X. et al., 2002) in the context of visual evoked potentials. For F_0, the noise was calculated from 100 to 200 Hz, excluding 100 and 200 Hz and the harmonics of 60 Hz (120 and 180 Hz) in that range. The reason for starting the estimation of the noise at 100 Hz is to exclude non-auditory related brain activity (e.g. gamma wave effects) on the responses from 25-100 Hz and low-frequency EEG noise. For F_1, the interval for estimating the surrounding noise ranged from 600 to 800 Hz excluding 600, 660, 700, 720, 780 and 800 Hz. Calculations of the rms-SNR at F_0 and F_1 were done over the entire steady-state response starting at 20 ms after the onset of the response. For each condition (the three stimulus variants in quiet and noise), the mean subject local rms-SNR, the standard deviation, and the standard error of the mean were also computed.

For F_0 and F_1 the local rms-SNR was calculated as:

$$rmsSNR = \sqrt{\text{signal power}} / \sqrt{\left(\sum \text{noise bins amplitude}^2\right) / \text{N}}$$

signal power being: $amplitude\_at\_target^2$ and N being the number of bins in the defined noise interval.

The responses at F_0 and F_1 were expected to increase in amplitude, and to have a higher local rms-SNR as we increase the stimulus SNR (Johnson et al., 2005; Song et al., 2006).

The fourth part of the analysis focused on determining whether or not the response at F_0 was non-linear relative to the contributions of the resolved and unresolved harmonics in the stimulus. To that end, the response to the Allformants stimulus was compared with the Vector Sum of the responses with the F_1Only and F_2&F_3Only stimuli.

The final part of the analysis focused on estimating the latency of the response at F_0, which would indicate where it originates in the auditory pathway. Latency was calculated as the difference in time between the start of the steady-state part of the stimulus and the start of the steady-state response for each stimulus variant, for both quiet and noisy conditions. The method for calculating latency followed the one described in Akhoun et al., 2008 (Akhoun et al., 2008a). First, since the sampling frequency of the stimulus (48 kHz) was different from that of the evoked responses (3202 Hz) up-sampling of the responses was performed using spline interpolation and re-sampling at 48 kHz. Next, the first 20 ms were removed from both the stimulus (of 300 ms) and the response (319.8 ms) to keep only the steady-state portion in the responses and the same starting point in both signals. Then, the envelope of the stimulus was obtained by passing the signal through a half-wave rectifier and a low-pass filter (with 501 filter coefficients and a cutoff frequency of 500 Hz) with compensation for the delay of the FIR filter incorporated into the analysis. The estimate of the latency was obtained by finding the time shift that maximizes the statistical correlation between the envelope of the stimulus signal and the response signal. The clean stimulus was used for evaluating the latency relative to the responses in both the quiet and noisy conditions. This is justified, because the source of the responses is based on the speech signal itself and not on the noise, the presence of which introduces uncertainties in the calculation of correlation, which would lead to inaccurate latency estimates. From each estimated latency obtained, 1 ms was subtracted from the estimated due to the delay in signal transmission up to the insert-earphones (Parbery-Clark et al., 2009).

3.7 Quality of Stimulus and Response Signals

Verification of the quality of the stimulus signal was an important step to perform before starting the actual recordings on the subjects. This was performed to ensure that the underlying speech signal in the stimulus remained at the same level when the noise was

added. Verification was made on the signal in the stimulus sound file, and also at the output of the sound generating system. Moreover, testing was done to verify that the acoustic system did not significantly alter the spectral shape of the stimuli. This included spectral analysis to look for any changes in position of the formants and their relative amplitudes. The measurement and analysis of the signal at the output of the sound generating system are discussed in Appendix B.

3.7.1 Artifacts, Electromagnetic Leakage, 60Hz-noise

Artifacts are unavoidable when recording auditory evoked responses. They originate from various sources such as the recording equipment or from non-auditory physiological activity such as tensed muscle, and they can create electric potentials that contaminate the auditory evoked response. When artifacts are generated, abnormally large potentials are observed in the auditory evoked response which affect not only the timing of the crucial elements forming the ABR, but also the signal to noise ratios which are important when analysing speech evoked potentials (Don and Elberling, 1994; Hall, 2007b; Akhoun et al., 2008a; Akhoun et al., 2008b). Moreover, the steady-state response is vulnerable to both movement artifacts and the state of wakefulness because they create a certain amount of noise in the response that could potentially lead to invalid results (Hall, 2007a).

The manufacturer of BioMARK™ included eighteen recommendations regarding the environmental electromagnetic noise reduction that should be followed before using the equipment. The list considered any potential sources in a typical shielded-soundproof booth that could affect the readings. The list included recommendations such as turning off fluorescent lights or dimmer switches when operating the equipment, and making sure all the devices are connected with the iso-transformer provided with the system. Since a system such as BioMARK™ uses electrodes to measure neurophysiological signals it is important to include an isolated barrier between the equipment and the power line so that no leakage currents (in microamps) cause microshock hazards to the subject (Webster and Clark, 1997). The noise reduction steps provided by the manufacturer were followed when possible, to ensure the quality of the acquired neuro-physiological signal.

One way to reduce any leakage from the sound generating equipment to the recorded response is to use alternating polarities for the recordings where the summation of the ABR with the positive and negative polarities of the speech stimulus together will create a cancellation of the effects of the artifacts (Hall, 2007b; Akhoun et al., 2008b). Another is to use insert-foam earphones connected to the speaker through plastic tubing instead of headphones since they generate a greater electromagnetic leakage (Akhoun et al., 2008b). A third option is to put as much distance as possible between the recording electrodes, and the earphone speakers when possible (Hall, 2007b). With insert earphones, the plastic tubing connecting the acoustic speakers to these earphones allows a greater distance to the electrodes.

There is a theoretical argument against the possibility of electromagnetic leakage in the EFR response at F_0. The connection between the sound generating equipment and the recording electrodes is usually modelled as a linear system (Picton and John, 2004; Akhoun et al., 2008b). Although the speech stimulus signal is modulated at the fundamental frequency, there is little energy present in the stimulus spectrum at F_0. Therefore, the generating and recording system cannot demodulate the stimulus signal and leak electromagnetic energy at F_0 from the stimulus to the response. This is also why the Auditory Steady-state Response (ASSR) at the modulation frequency is considered immune to electromagnetic leakage (Picton and John, 2004; Akhoun et al., 2008b). It is important to note that this argument only applies to the EFR, and not the FFR, because this response is due to the demodulation of the envelope and not to a specific frequency component in the stimulus as it is the case for the FFR.

We experimentally tested for the presence of electromagnetic leakage before starting the ABR recordings, in order to verify that the BioMARK™ system does not introduce electric potentials in the collected response. To perform this test, electrodes were placed on the subject's scalp, and in all other aspects the experiments proceeded as normal, with the sole exception that instead of having the insert-earphone put in the subject's ear, it was inserted into a G.R.A.S. 2cc coupler (IEC 711, RA0045) connected to a turned-off SLM. The role of the coupler is to act as a substitute for the subject's ear, since it presents approximately the same acoustic load to the transducer as if the earphones were inserted in the subject's ear.

Since the subject must not hear any acoustic sound, earplugs were inserted in the subject's ear canal. If there is no electromagnetic leakage, the Fourier analysis performed on the response collected from the electrodes would show a spectrum in which harmonic components of the stimulus are absent. If leakage contaminates the evoked response, additional techniques to eliminate the leakage in the recordings need to be performed prior to the analysis (Don and Elberling, 1994; Leonardo et al., 2005; Sanchez and Gans, 2006; Hall, 2007b). Here, the spectrum of the average of the alternating polarity recordings confirmed the absence of electromagnetism leakage in the EFR (Appendix B, figure B-4). A similar analysis on the FFR obtained with the coupler showed no electromagnetic leakage at F_1.

When performing the test for electromagnetism leakage, a major power-line contamination artifact causing significant peaks to appear at 60 Hz in the frequency spectrum of the ABRs was discovered (see figure 3-14). Loops in an extension cord and a computer screen being too close to the iso-transformer inside the sound booth were found to be the source of contamination. After making necessary arrangements of the equipment and the cables, the same test was repeated and the 60 Hz peak was considerably minimized in the frequency spectrum of the ABRs (see figure 3-15). It should be noted, however, that suppression of power-line contamination was not successful in all recordings.

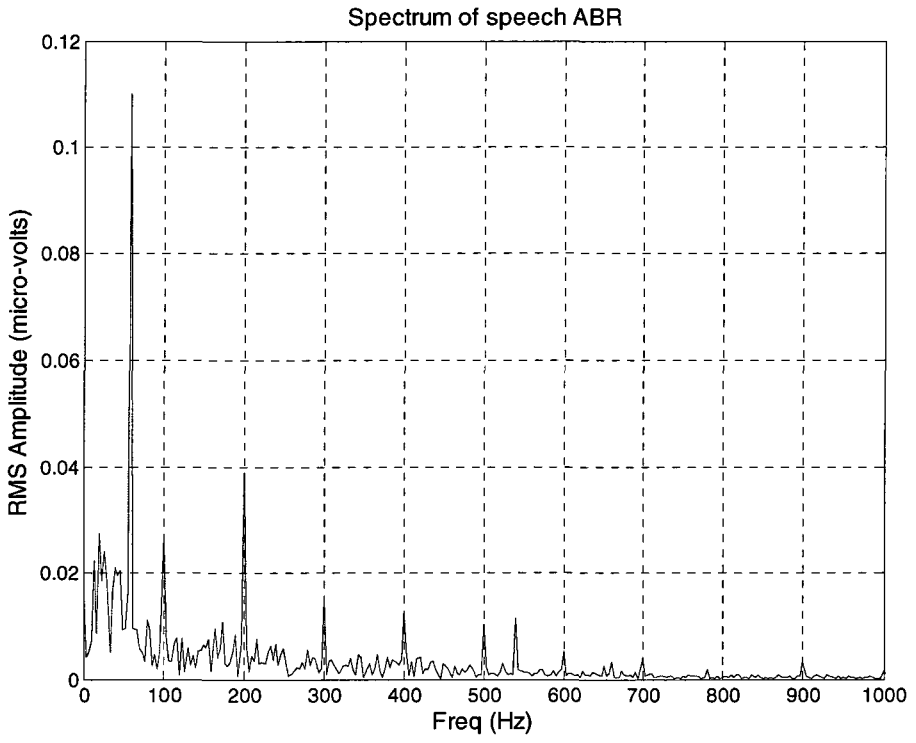

Figure 3- 14: Frequency spectrum of an ABR (obtained using a clean stimulus) during electromagnetic leakage test showing artifacts at the power-line frequency of 60Hz and at 530Hz (from an unknown source)

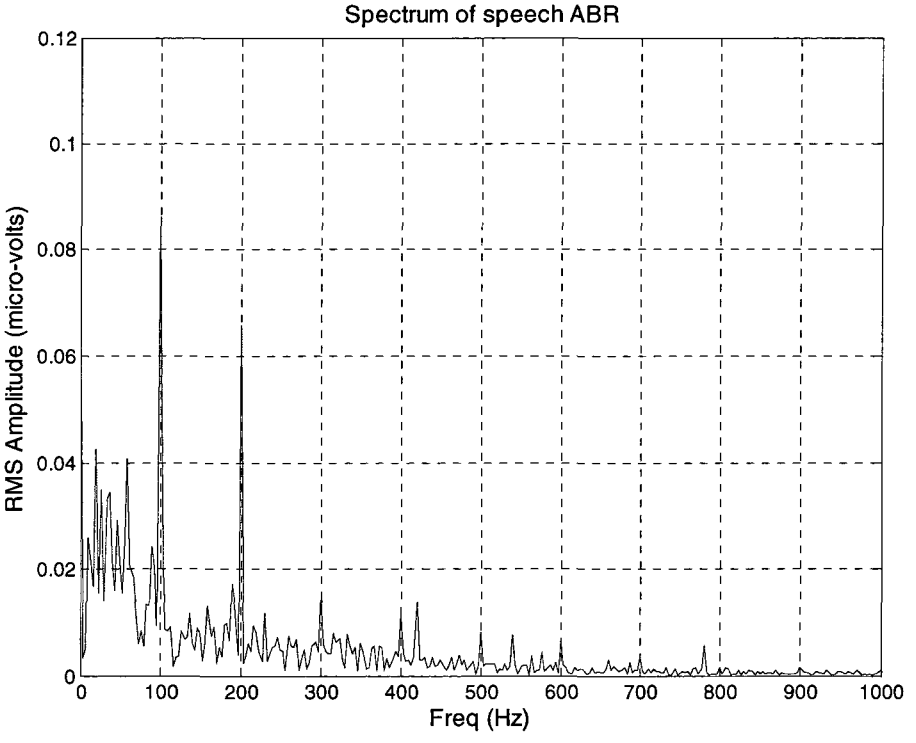

Figure 3- 15: Frequency spectrum of the ABR (second attempt)

Power-line contamination could potentially affect the evaluation of the transient response, since analysis is done on it in the time domain. Since the fundamental frequency is located at 100 Hz, it made it difficult to use a notch filter to eliminate components at 60 and 120 Hz. Therefore, minimizing the level of power-line contamination was the preferable solution. On the other hand, the steady-state responses at F_0 (100 Hz) and at F_1 (700 Hz) are not susceptible to power-line noise, because they are not harmonics of 60 Hz.

3.8 Statistical Analysis

Because the stimulus signal had different variants (Allformants, F_1Only, and F_2&F_3Only) and its SNR also varied (i.e. SNR:-5 and quiet), two-way repeated measures ANOVA (more than one test done on each subject) had to be employed.

The steady-state response was investigated by performing statistical analysis on the amplitudes and local SNR at F_0 and at F_1, with a significance level (α) set at 0.05. On the other hand, the transient response was studied by looking at the effect of noise on the slope of the VA onset, and the amplitude of the onset peaks V and A in the grand-averaged signal, which have been reported to be affected by noise (Kraus et al., 1995a; Kraus and Cheour, 2000; Russo et al., 2004). No ANOVA was performed on the transient responses because responses from individual subjects were too noisy.

Before comparisons were made between the different conditions using the ANOVA, we needed to confirm that, at least in most recordings, the spectral components at the target frequencies are not merely noise. Therefore, we used a statistical test that has been applied to these types of measurements to determine that the signal at F_0 and F_1 is not noise at a given confidence level ($p<0.05$). In this test, an F-ratio is computed based on the strength of the signal at the target frequency versus the strength of the adjacent noise frequency bins (Dajani et al., 2005; Skoe and Kraus, 2010). For F_0 at 100 Hz, the noise bins ranged from 100-200 Hz excluding 100, 120, 180 and 200 Hz, while for F_1 (700 Hz), the noise bins ranged from 600-800 Hz, excluding 600, 660, 700, 720, 780 and 800 Hz. The p-values and the F-ratios were computed using a pre-existing program made using LabVIEW™ version 8.6 (National

Instruments Inc.). For example if $p<0.05$, then we concluded that there was a 95% probability that the signal at the target frequency was different from the surrounding noise.

The ANOVA was done using the SPSS software PASW Statistics 17 (SPSS Inc., Chicago, IL) with the background condition (quiet and noise) and stimulus content (Allformants, F_1Only, and F_2&F_3Only) set as independent variables. The amplitude and local SNR of the responses at F_0 and F_1 were the dependent variables. Next, there was a concern if sphericity was present in the collected data. Sphericity is a measure of how much the data is correlated. The significance level obtained from Mauchly's test confirms if sphericity can be assumed or not (Field, 2009). If the probability value or confidence level (i.e. p-value) obtained from the Mauchly's test is greater than 0.05, then sphericity is assumed. Sphericity must be considered when there are three or more repeated-measures conditions. Where sphericity is not assumed, reliance is made on the significance level given by the Geenhouse-Geisser correction.

Post-hoc tests

The ANOVA on the amplitudes and the local SNRs of the responses at F_0 and F_1 gave statistically significant differences between the means in some cases. To discover which groups were different from the others, we performed multiple pair-wise post-hoc comparisons between the groups using repeated measures two-tailed t-tests. Bonferroni corrections were applied when required to prevent 'Type I' errors, i.e. that the means are statistically different when in fact they are not (Norman and Streiner, 2008). This step was completed using the SPSS software as well.

The Fourier analysis performed on the steady-state response at F_0 gives a vector having a magnitude and a phase. Comparing the vectors in terms of magnitude and phase helped us determine the contribution of the resolved and unresolved harmonic regions, due to the F_1Only and F_2&F_3Only stimuli respectively. In addition, comparing the responses at F_0 generated using the Allformants stimuli versus the Vector Sum of the responses of the other two variants (F_1Only and F_2&F_3Only) would indicate whether or not the responses to the two regions add linearly. The response at F_1 was not considered because one of the three stimulus

variants (the F_2&F_3Only stimulus) does not contain strong components at F_1 and so does not produce a response at F_1 (based on the principle of the FFR). Repeated measures two-tailed t-tests were applied to see if the magnitudes of the vectors (Allformants vs. Vector Sum) for both the quiet and noisy conditions were statistically different.

Chapter 4 : Results

The results obtained from the analysis performed on the auditory evoked responses are presented in this chapter. One aspect of the analysis consisted in investigating the effect of noise on the transient and sustained responses at F_0 and F_1 produced by the different stimulus variants. This study also looked at the contribution of the resolved and unresolved harmonics to the response at F_0 and investigated whether the responses from the resolved and unresolved harmonics combined linearly in the overall response. The results presented here also provide an insight on the origin of the response at F_0 using latency analysis.

4.1 Time domain grand-averaged responses

Figures 4-1 to 4-6 show the time domain grand-average responses specific to the EFR and figures 4-7 to 4-12 to the FFR for the three stimulus variants for quiet and noisy conditions. These responses were averaged over twelve subjects with an epoch recording time of 319.8 ms. Note that since the stimulus is 300 ms, the extra 19.8 ms at the end were removed.

4.1.1 Envelope Frequency Response

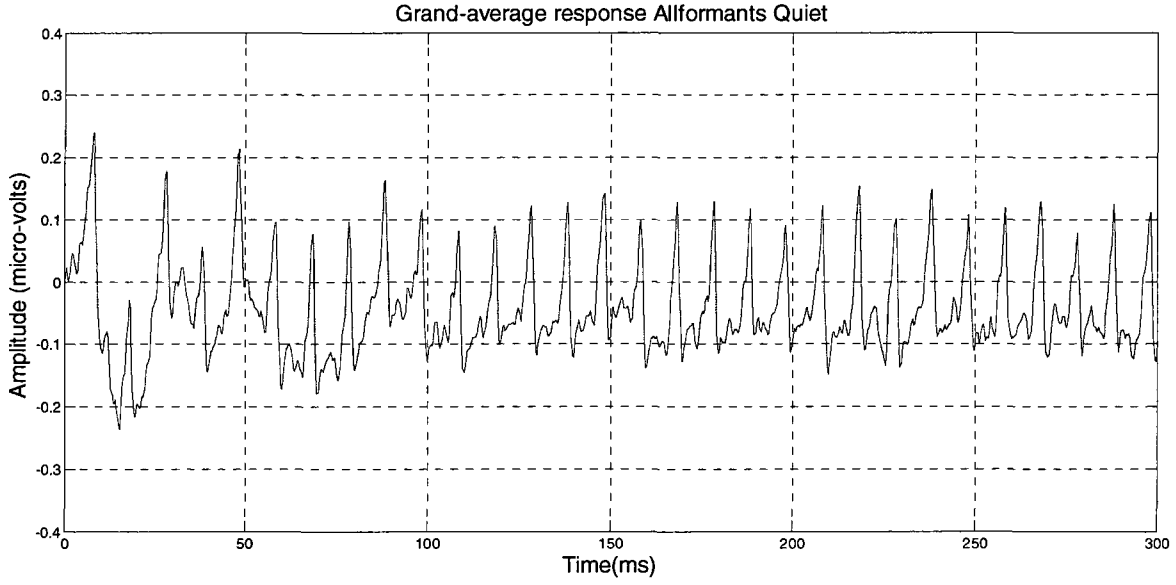

Figure 4-1: Grand-average response Allformants, Quiet (EFR)

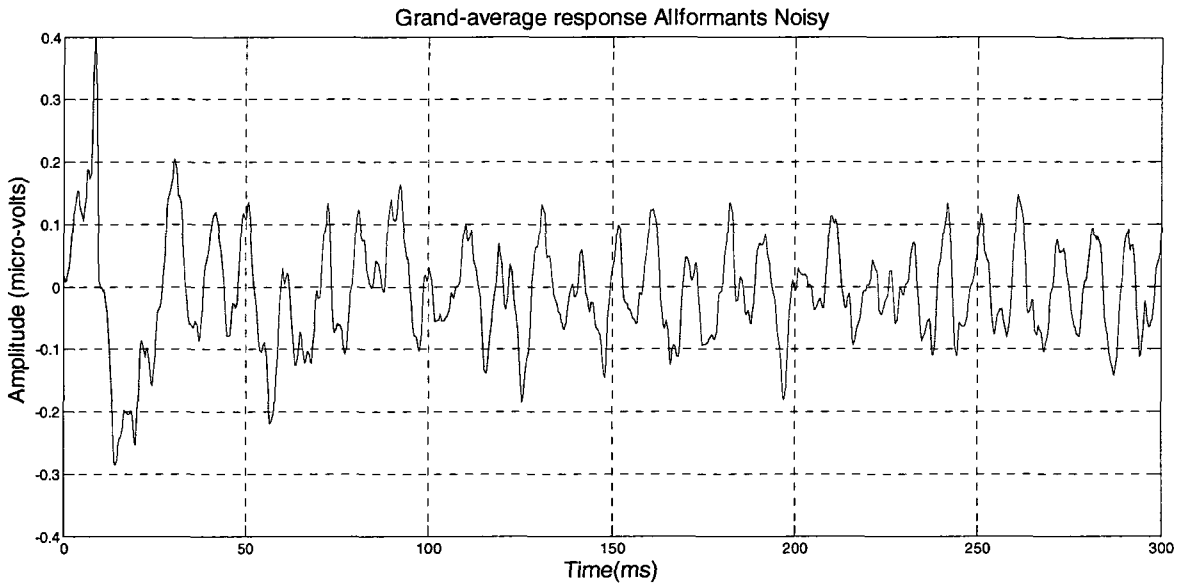

Figure 4-2: Grand-average response Allformants, Noisy (EFR)

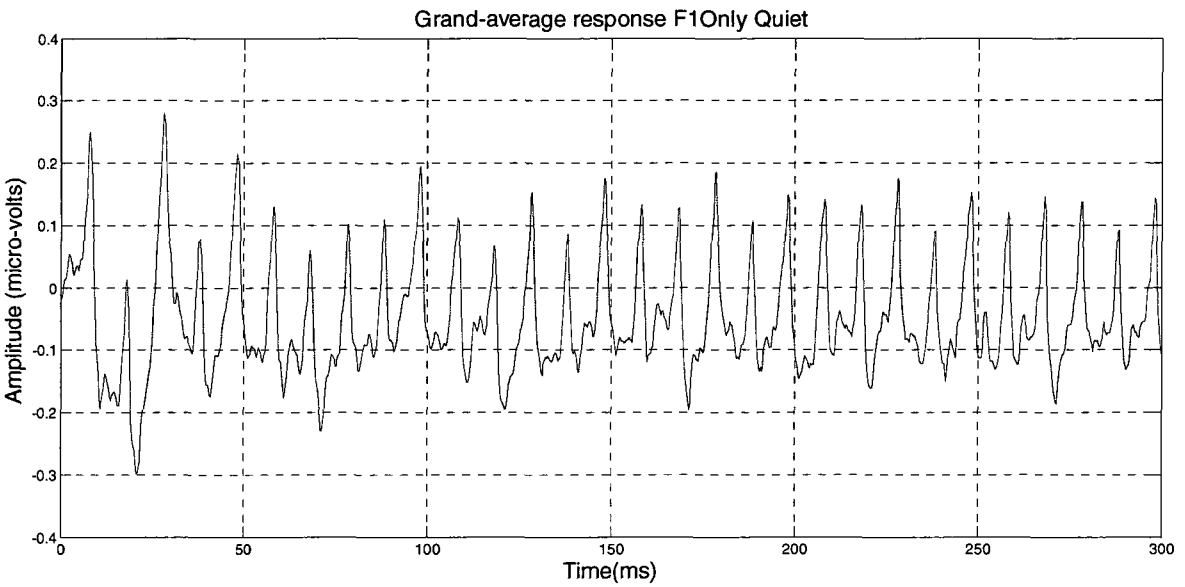

Figure 4-3: Grand-average response F_1Only, Quiet (EFR)

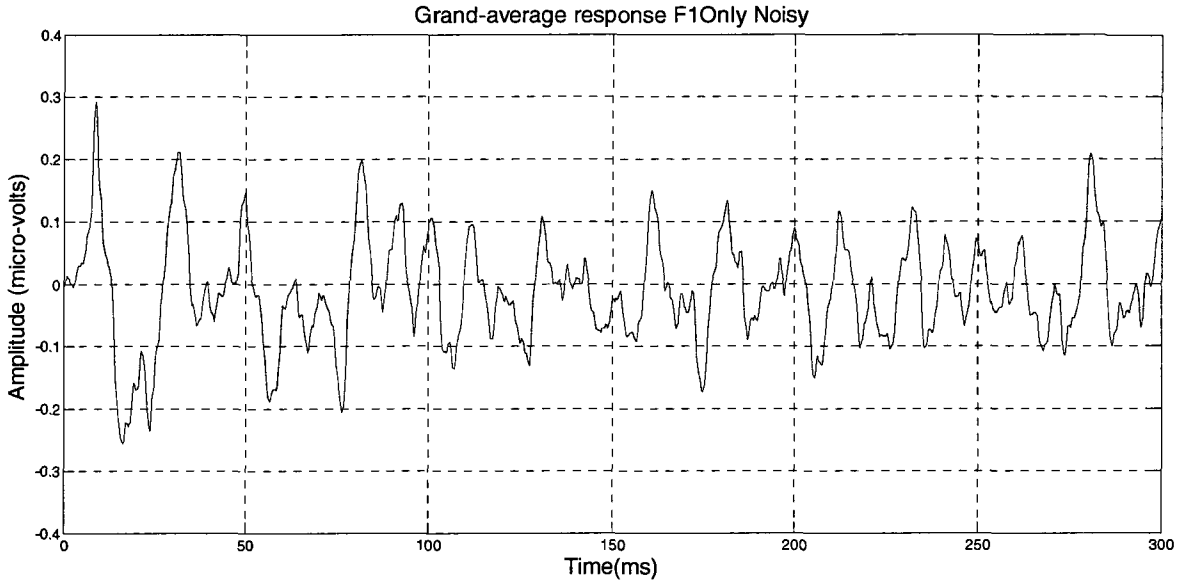

Figure 4-4: Grand-average response F_1Only, Noisy (EFR)

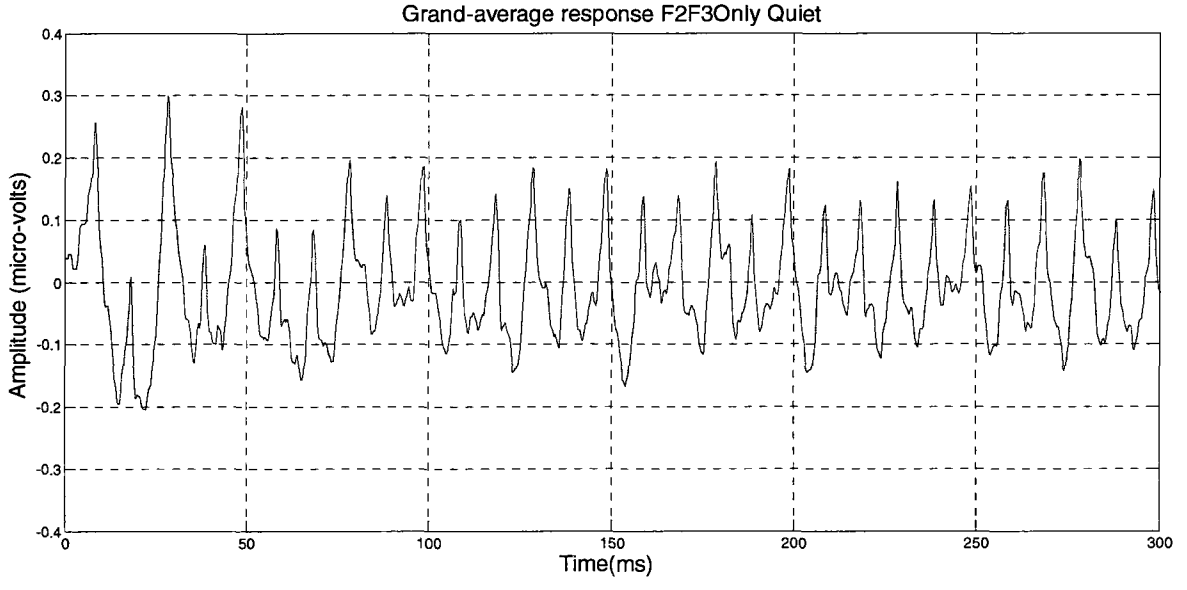

Figure 4-5: Grand-average response F_2&F_3Only, Quiet (EFR)

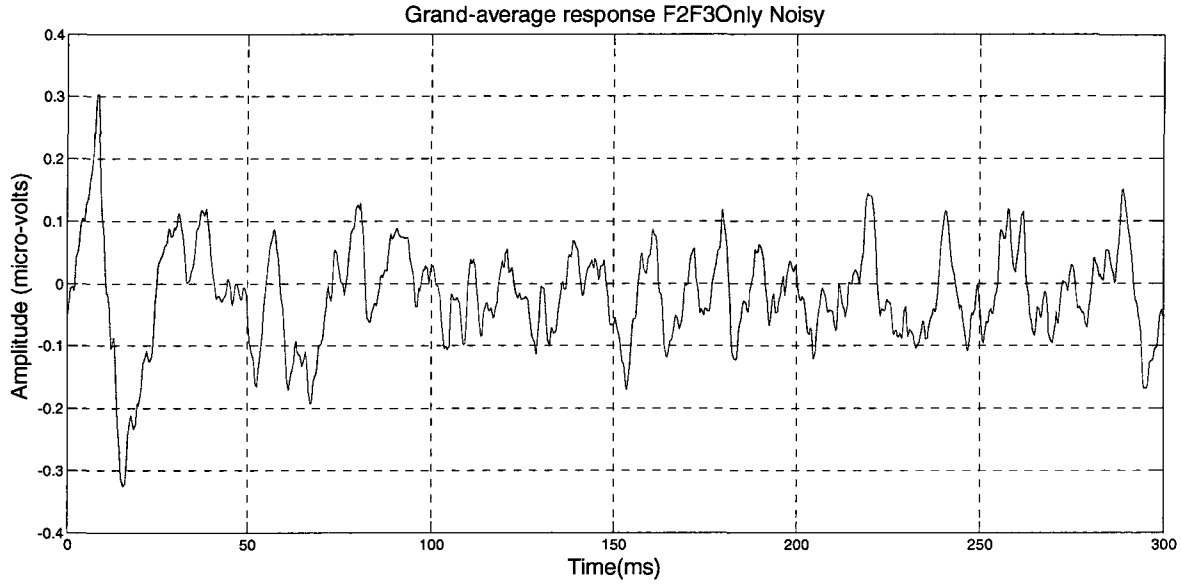

Figure 4-6: Grand-average response F_2&F_3Only, Noisy (EFR)

4.1.2 Frequency Following Response

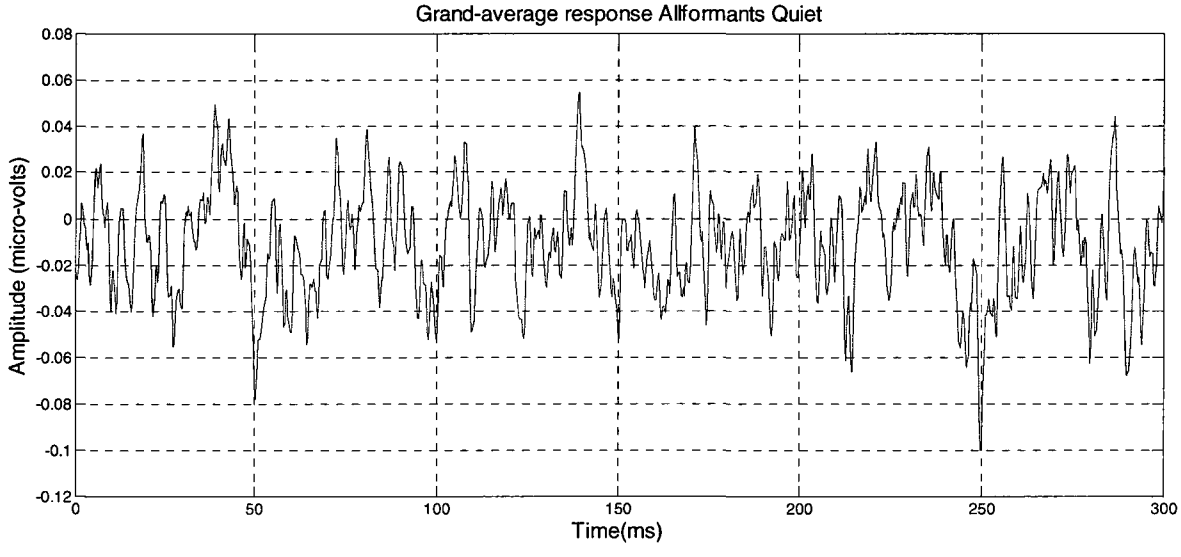

Figure 4-7: Grand-average response Allformants, Quiet (FFR)

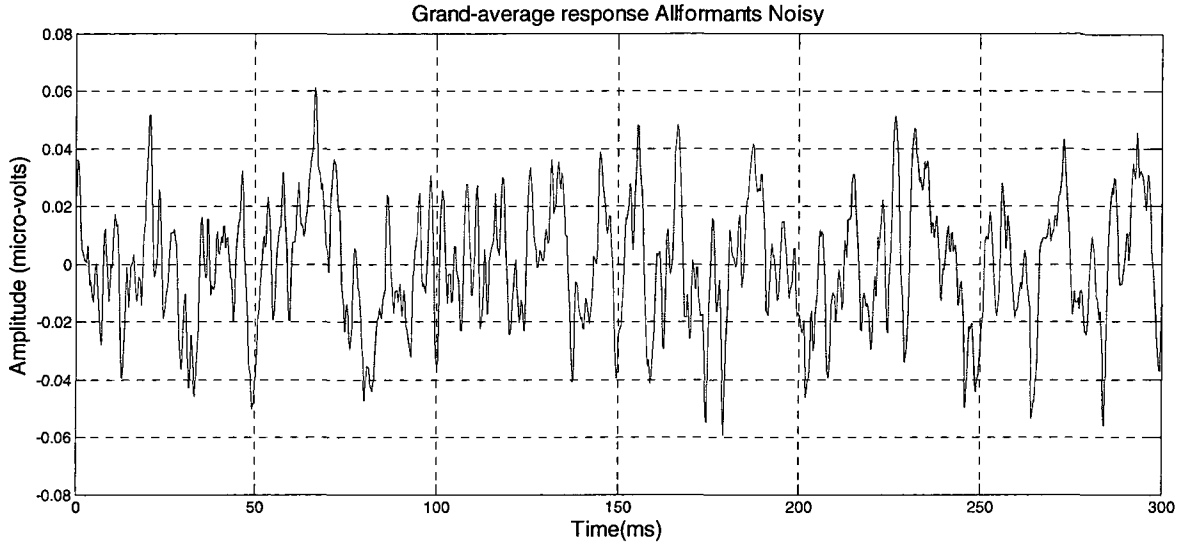

Figure 4-8: Grand-average response Allformants, Noisy (FFR)

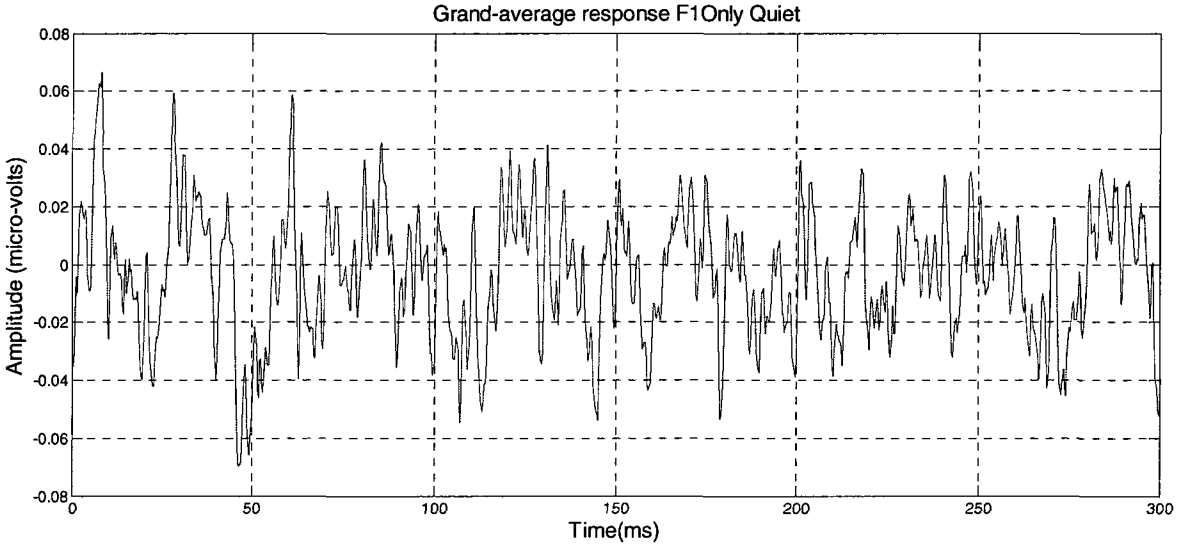

Figure 4-9: Grand-average response F_1Only, Quiet (FFR)

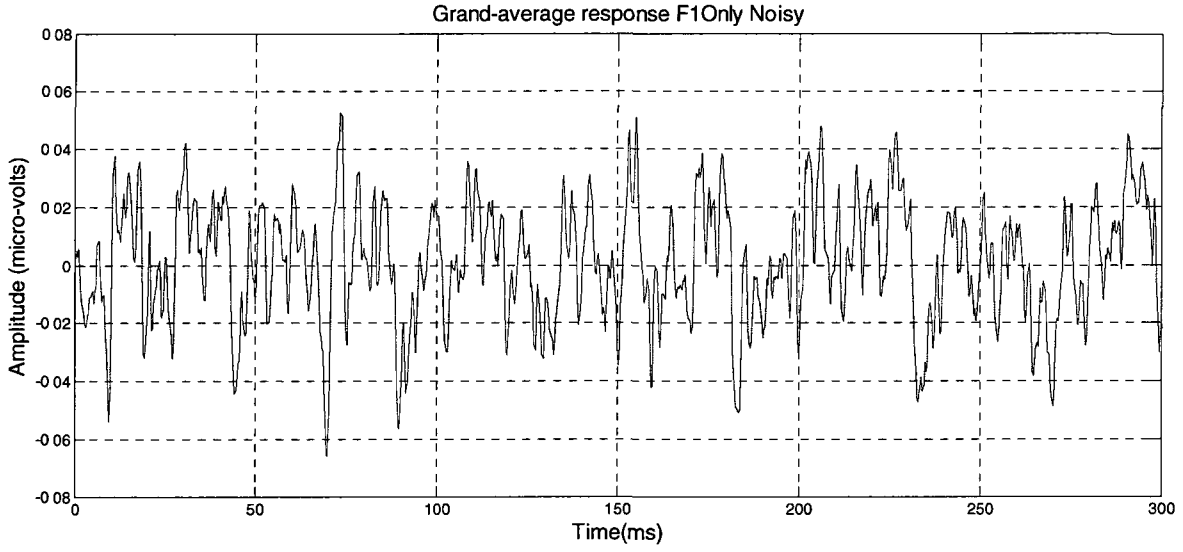

Figure 4-10: Grand-average response F_1Only, Noisy (FFR)

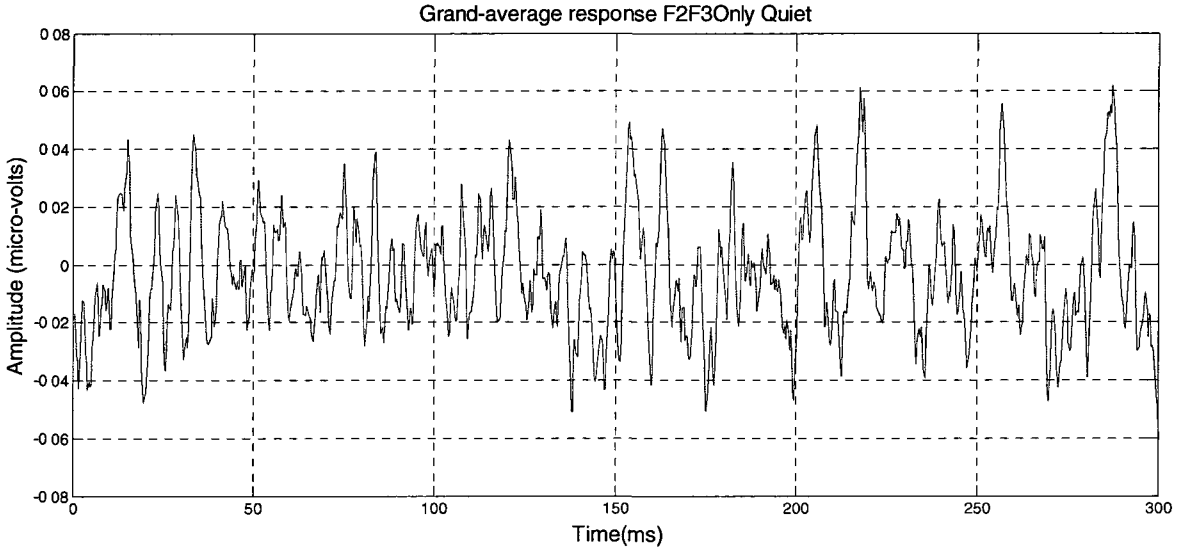

Figure 4-11: Grand-average response F_2&F_3Only, Quiet (FFR)

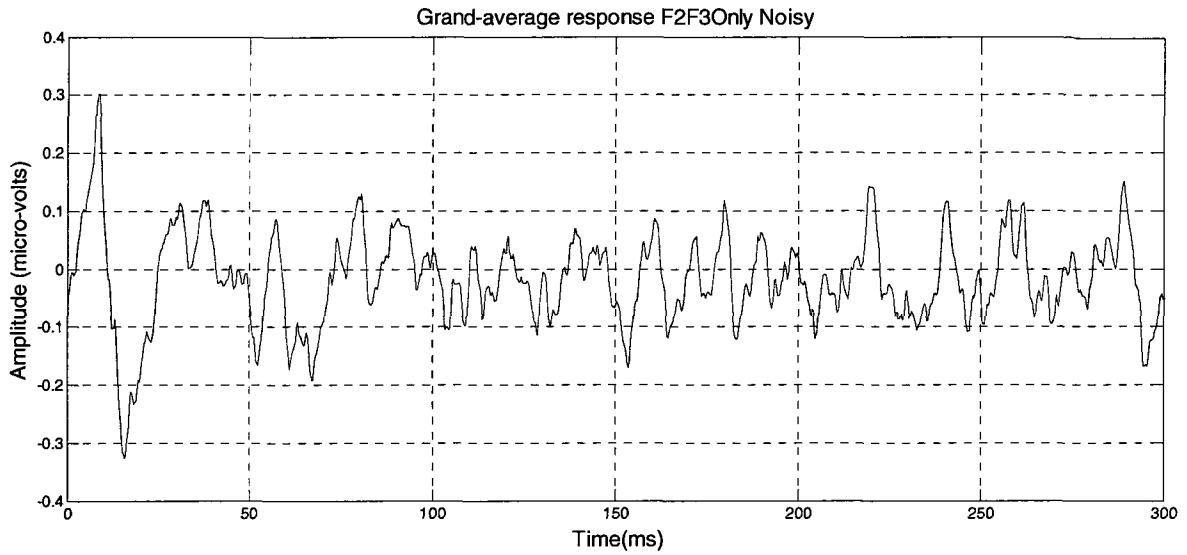

Figure 4-12: Grand-average response F_2&F_3Only, Noisy (FFR)

4.2 Analysis of the Transient Response

The interest we had in analyzing the transient response evoked using a quiet and noisy stimulus derives from various studies showing an effect of background noise on the VA onset complex (Russo et al., 2004; Johnson et al., 2005; Kraus and Nicol, 2005; Song et al., 2006). Using the grand-averaged responses to both the clean and the noisy stimulus (in the +- configuration), we examined the location proposed in the literature for the VA complex in the region from 8-12 ms. An expert in clinical auditory electrophysiology helped us identify the peaks V and A and the VA slope in the averaged transient response. They are shown in the figures 4-13 and 4-14 for both quiet and noisy conditions.

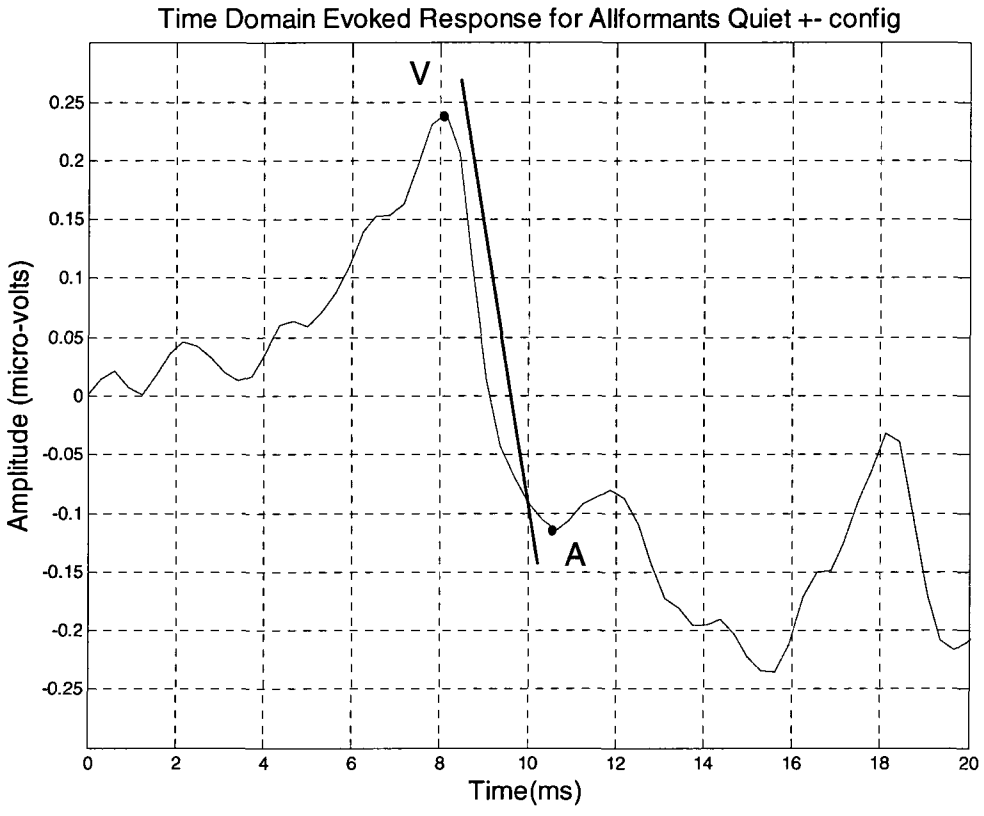

Figure 4-13: Identification of the VA complex from the clean stimulus' response

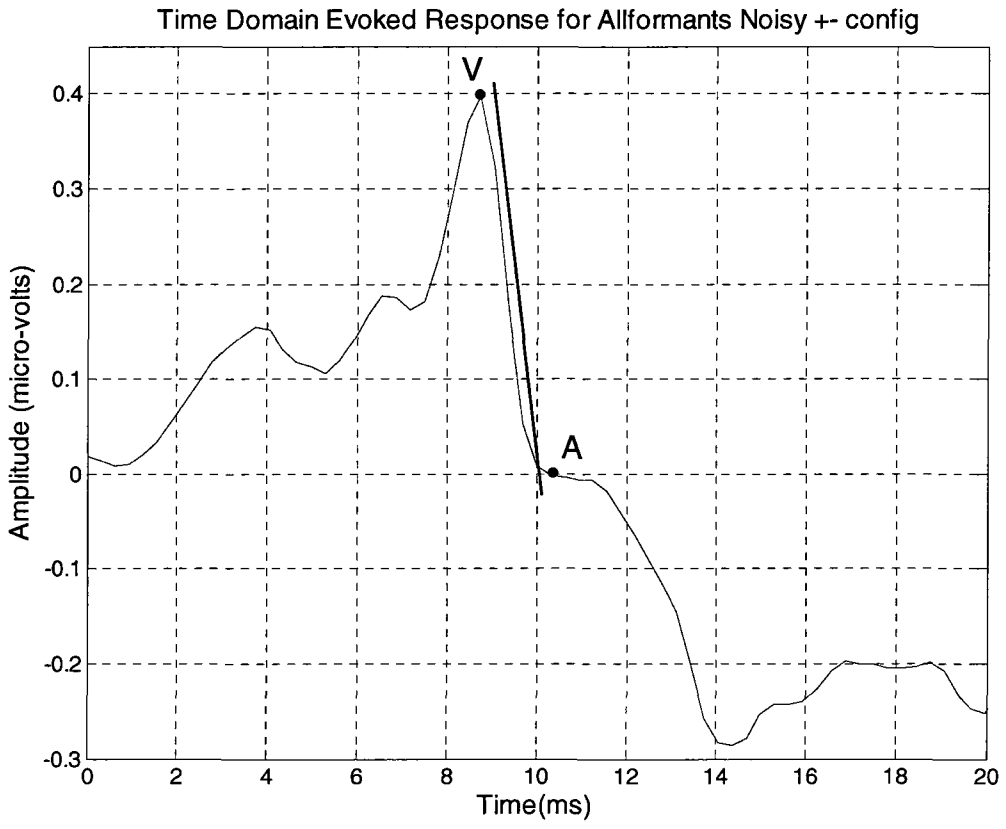

Figure 4-14: Identification of the VA complex from the noisy stimulus' response

To approximate the slope of the VA complex we proceeded by drawing a line following the drop in the waveform between V and A. The slope of the VA onset for the quiet condition is approximately -0.25 μV/ms and for the noisy condition is approximately -0.40 μV/ms. It also appears that the addition of noise induced a delay of 0.5-1 ms for the VA onset, and the amplitude of the V wave appears higher.

4.3 Analysis of the Sustained Response

Detailed numerical results are tabulated in Appendix C, D, E and F. The following sections provide a summary.

4.3.1 Detection of the Steady-state Response at F_0 and F_1

With the Allformants stimuli, most of the responses at F_0 (in the EFR) were statistically higher than the surrounding noise (11 out of 12 subjects) for both quiet and noisy conditions (see Appendix E, tables E-1 and E-2). Similar results were obtained for the F_1Only variant (11 subjects for the quiet and 12 for the noisy conditions). For the F_2&F_3Only variant, most of the responses were statistically above the surrounding noise for the quiet condition (12 subjects) and for the noisy condition (10 subjects).

With the Allformants stimuli, most of the responses at F_1 (in the FFR) were statistically higher than the surrounding noise (11 subjects for the quiet condition and 10 subjects for the noisy condition) (see Appendix E, tables E-3 and E-4). For F_1Only, similar results were obtained (11 subjects for the quiet condition and 9 subjects for the noisy condition). For F_2&F_3Only, the results show that the responses at F_1 were generally not statistically above noise (4 subjects for the quiet condition and none for the noisy condition).

The same operations were done on the grand-averaged responses where the response at F_0 was revealed to be statistically stronger than noise for all three variants (Allformants, F_1Only and F_2&F_3Only) for both quiet and noisy conditions. For F_1, the responses were statistically higher than noise for all three variants for the quiet condition. For the noisy condition only the grand-averaged responses from the Allformants and F_1Only variants were statistically above noise.

4.3.2 The Envelope Frequency Response (EFR) at F_0

The time and frequency-domain grand-averaged Envelope Frequency Responses are shown in figures 4-15 and 4-16 respectively. The ANOVA on the amplitude of the Envelope Frequency Responses at F_0 indicates that there is no major effect resulting from adding noise to the stimulus ($F_{(1,11)}$=2.625, p=0.134), but there is a significant effect related to the content of the stimulus ($F_{(2,22)}$=5.503, p<0.05), and the interaction between the type of stimulus (noisy or quiet) and its content ($F_{(2,22)}$=8.666, p<0.01). Table 4-1 provides a summary of the amplitudes of the individual responses for the stimulus variants (quiet and noisy condition).

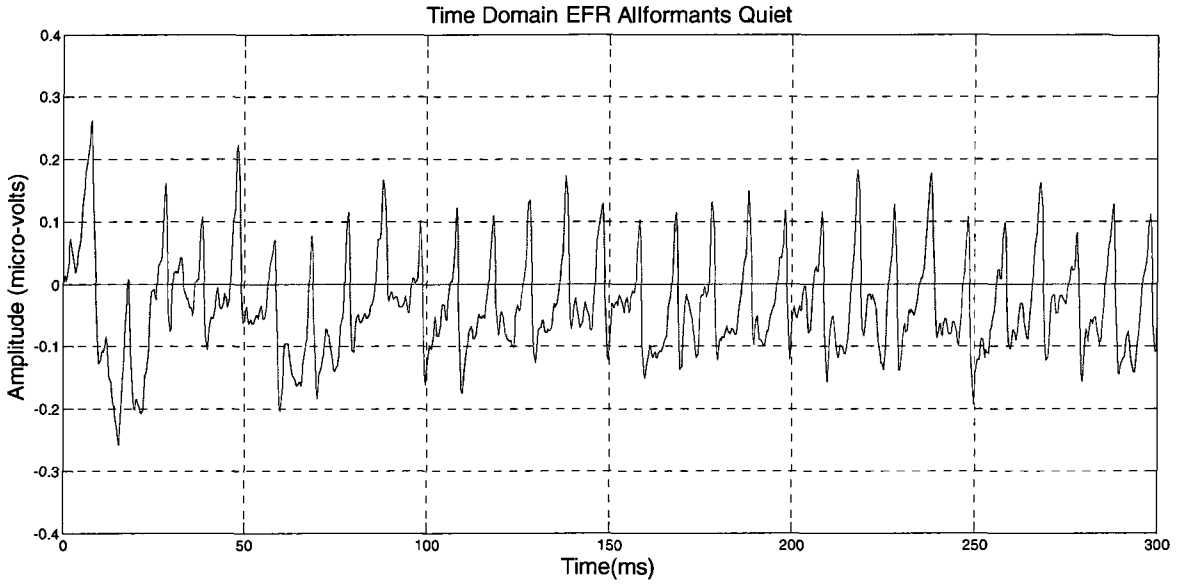

Figure 4-15: Grand-average time domain EFR (Allformants, quiet)

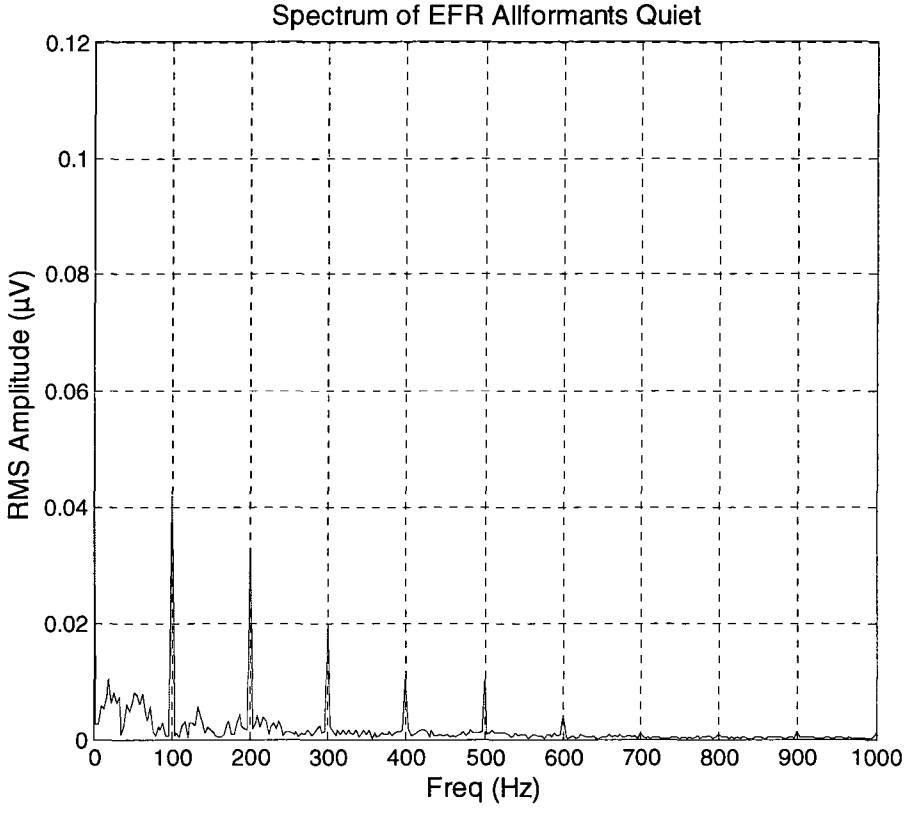

Figure 4-16: Grand-average frequency spectrum of EFR (Allformants, quiet)

rms amplitude (μV)	Allformants		F_1Only		F_2&F_3Only	
	Quiet	Noisy	Quiet	Noisy	Quiet	Noisy
Minimum	0.007096	0.004287	0.008886	0.013467	0.021545	0.004077
Maximum	0.130035	0.139898	0.143111	0.122962	0.135919	0.058723
Mean	0.05896	0.06189	0.06919	0.05505	0.06831	0.02718

Table 4- 1: Summary of individual responses amplitude for EFR at F_0

A similar analysis was performed using the local rms-SNRs at F_0. The results show that adding noise to the stimulus had an impact on the response at F_0 ($F_{(1,11)}$=6.891, p<0.05). The variation in content of the stimulus also had an effect ($F_{(2,22)}$=6.174, p<0.01), and the interaction between type (noisy or quiet) and content was significant ($F_{(2,22)}$=3.846, p<0.05). The detailed results of the amplitudes and rms-SNRs for individual subjects response at F_0 can be found in Appendix C. Table 4-2 provide a summary of the rms-SNRs obtained for the individual responses.

rms SNR	Allformants		F_1Only		F_2&F_3Only	
	Quiet	Noisy	Quiet	Noisy	Quiet	Noisy
Minimum	0.902867	0.494907	0.946596	1.877180	2.255210	0.430822
Maximum	18.907100	21.813800	24.492500	19.585700	31.354100	5.372940
Mean	9.09195	8.39244	12.02184	7.57619	10.78810	3.29847

Table 4- 2: Summary of individual responses rms SNRs for EFR at F_0

The following bar graphs (figures 4-17 and 4-18) summarize the results of the post-hoc pair-wise comparisons for amplitude and rms-SNR respectively for the EFR responses at F_0 across the stimulus variants for both quiet and noisy conditions. The detailed results of the statistical analysis for the response at F_0 can be found in Appendix F (Tables F-1, F-2, F-5, and F-6).

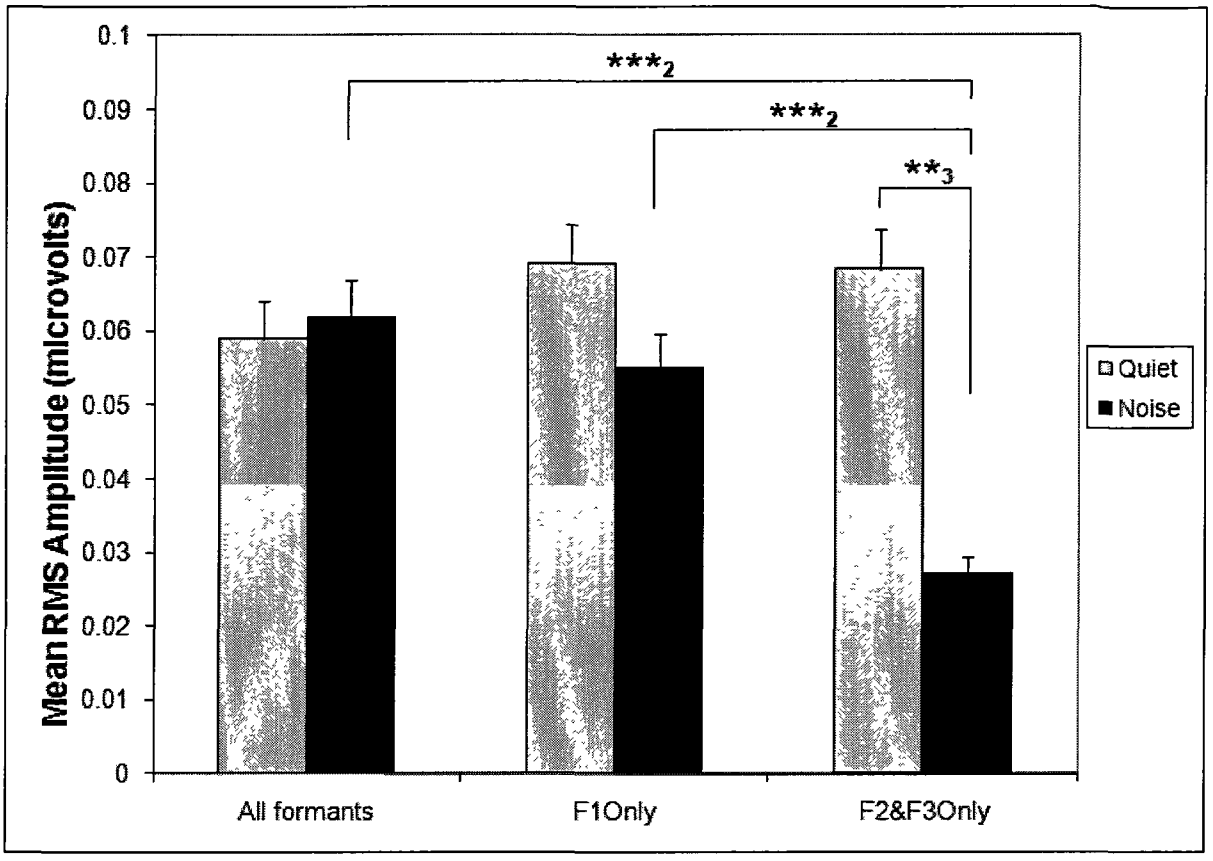

Figure 4-17: Comparison of the amplitudes of the responses at F_0 across the type of stimulus (quiet and noisy) and content (Allformants, F_1Only, and F_2&F_3Only)

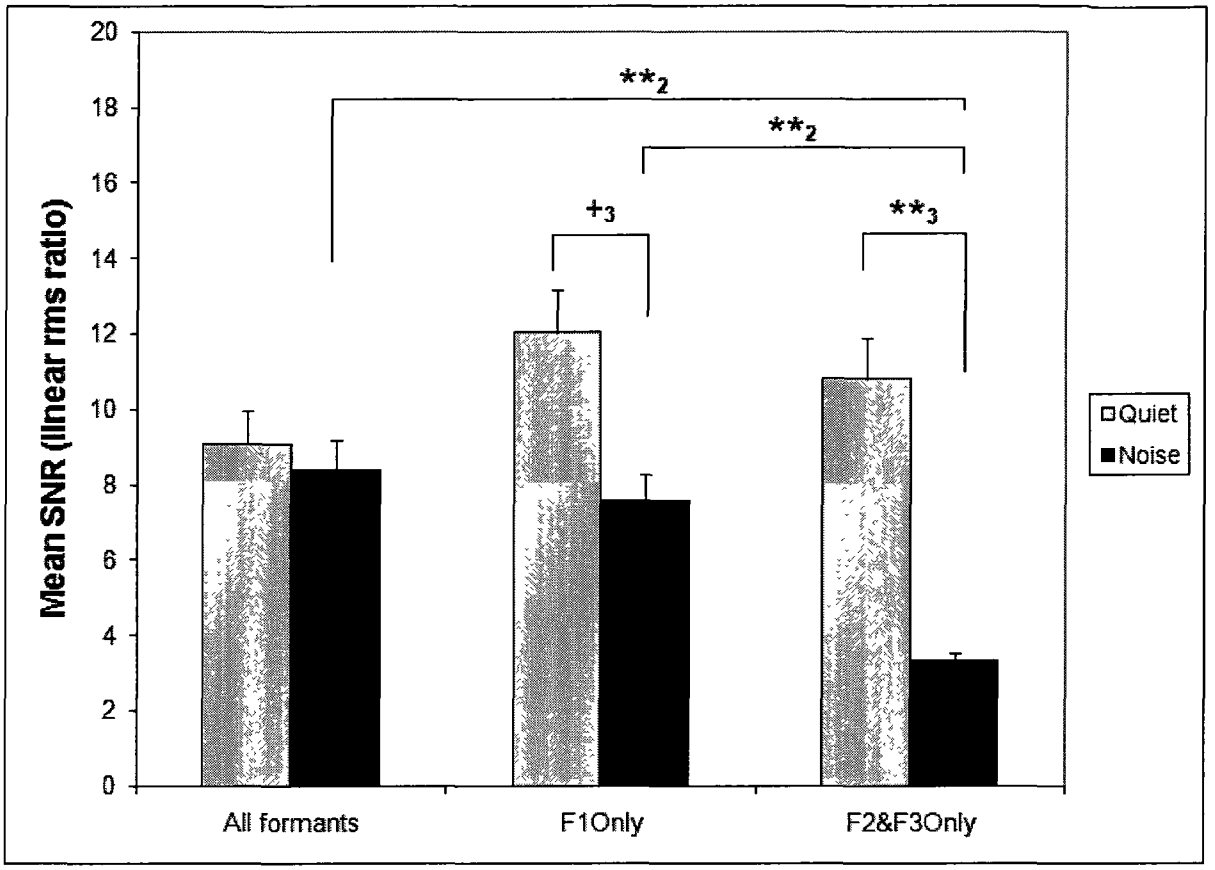

Figure 4-18: Comparison of the local rms-SNR of the responses at F_0 across the type of stimulus (quiet and noisy) and content (Allformants, F_1Only, and F_2&F_3Only)

The number of stars above the comparisons indicate the level of significance: $p<0.001$='***', $p<0.01$='**', $p<0.05$='*', and tendency towards significance: $0.05<p<0.1$= '+'. The bold number next to the stars refers to one of the following experimental questions that the comparison relates to:

1) When the signal is clean, do Allformants, F_1Only, F_2&F_3Only contribute differently to the response?
2) When the signal is noisy, do Allformants, F_1Only, F_2&F_3Only contribute differently to the response?
3) a) Does noise change the response with the Allformants stimulus?
   b) Does noise change the response with the F_1Only stimulus?
   c) Does noise change the response with the F_2&F_3Only stimulus?

For questions 1) and 2) the Bonferroni correction was applied, because they involved multiple comparisons. Therefore, the confidence level was corrected with respect to the number of comparisons made. Since we were comparing three means for each condition (Allformants, F_1Only or F_2&F_3Only) the p-value had to be less than 0.05/3 for it to be significant at the same confidence level. For questions 3a) b) and c) the comparisons were performed between each respective means for the quiet vs. noisy conditions for each of the three stimulus variants (i.e. F_1Only-Clean with F_1Only-Noisy, etc.). Therefore, for the third question no correction was applied because one comparison was made at a time.

The results show that for the response at F_0, there was no significant difference between the amplitudes for the various content conditions (Allformants, F_1Only, and F_2&F_3Only) of the quiet stimulus. Adding noise did not change the response amplitude at F_0 with the Allformants or the F_1Only stimulus. For the F_2&F_3Only stimulus in noise, the response at F_0 was significantly weaker compared with the Allformants and F_1Only stimuli ($p<0.001$).

The results on the rms-SNRs of the responses at F_0 support the results obtained with the amplitudes. For the quiet condition, there was no significant difference between the means with the three types of stimulus content. Adding the noise significantly reduced the rms-SNR with the F_2&F_3Only stimulus.

4.3.3 The Frequency Following Response (FFR) at F_1

The time and frequency-domain grand-averaged Frequency Following Responses are shown in figures 4-19 and 4-20 respectively. The ANOVA was also performed on the amplitude of the Frequency Following Responses at F_1. The Mauchly's test revealed that sphericity was not present in the interaction between the type of stimulus and its content ($p>0.05$). Therefore, the probability value *p* for the ANOVA was taken after the Greenhouse-Geisser correction. The detailed results of the amplitudes and rms-SNRs for individual subjects response at F_1 can be found in Appendix D. Tables 4-3 and 4-4 provide a summary of the amplitudes and rms-SNRs of the individual responses for the stimulus variants (quiet and noisy conditions).

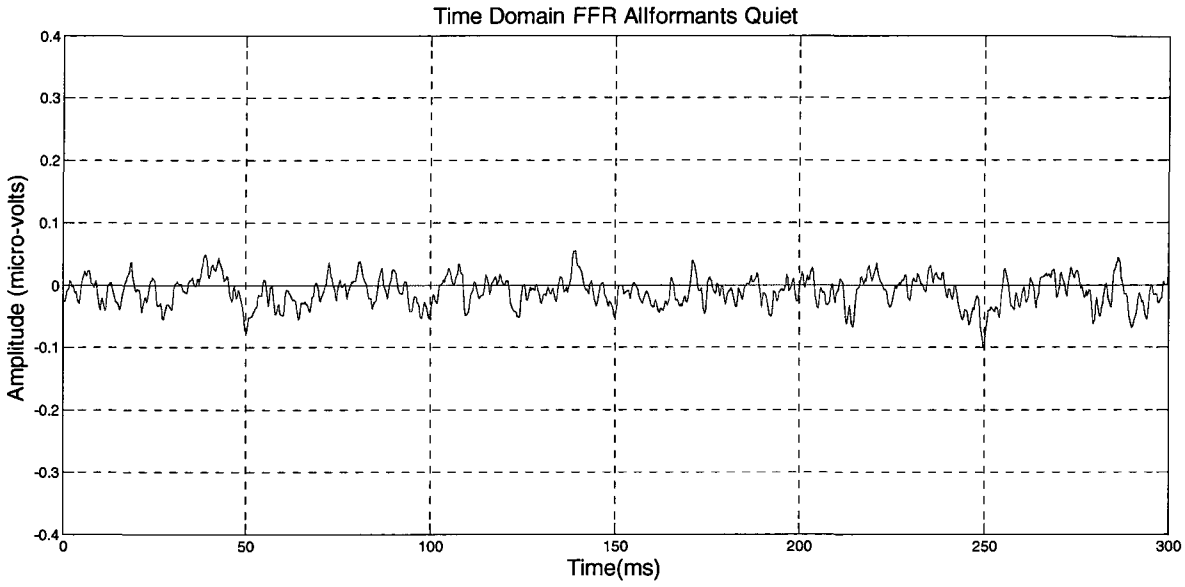

Figure 4-19: Grand-average time domain FFR (Allformants, quiet)

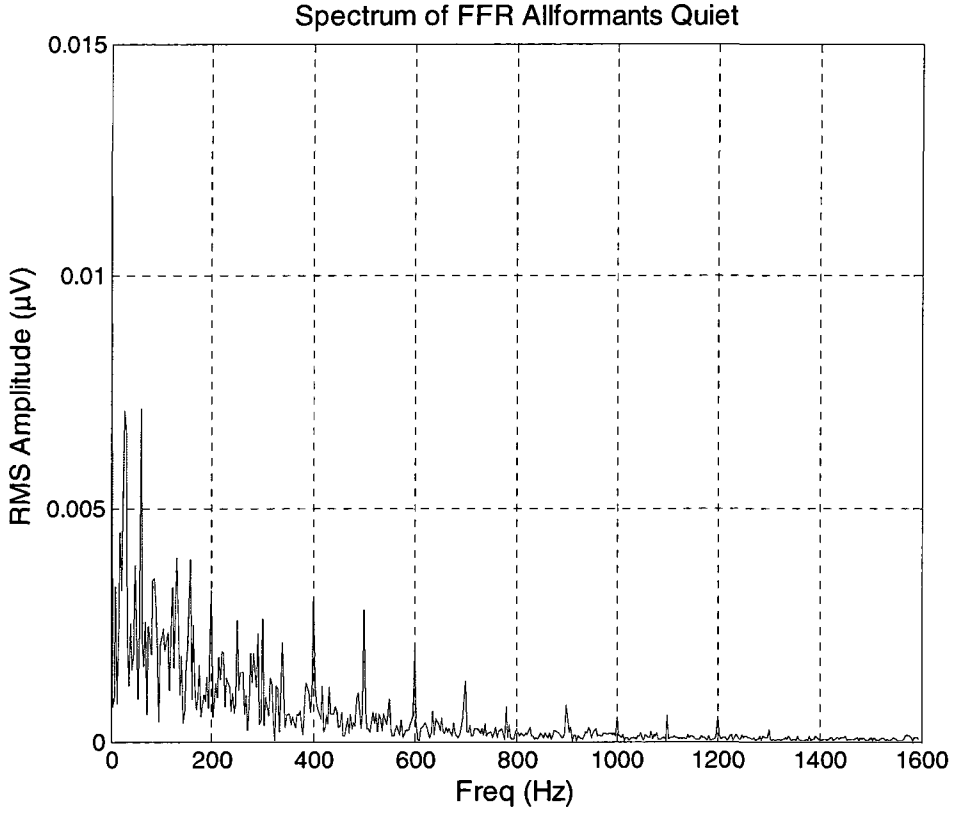

Figure 4-20: Grand-average frequency spectrum of FFR (Allformants, quiet)

rms amplitude (μV)	Allformants		F_1Only		F_2&F_3Only	
	Quiet	Noisy	Quiet	Noisy	Quiet	Noisy
Minimum	0.000916	0.001107	0.000916	0.000371	0.000110	0.000179
Maximum	0.008750	0.006236	0.011057	0.004938	0.002606	0.001244
Mean	0.00438	0.00293	0.00444	0.00280	0.00119	0.00078

Table 4- 3: Summary of individual responses amplitude for FFR at F_1

	Allformants		F_1Only		F_2&F_3Only	
rms SNR	Quiet	Noisy	Quiet	Noisy	Quiet	Noisy
Minimum	1.417150	0.001360	0.768534	0.729260	0.110974	0.243121
Maximum	15.364100	8.303820	15.414700	6.488850	2.791930	1.690450
Mean	5.87525	3.36734	6.17579	3.29302	1.56085	0.88009

Table 4- 4: Summary of individual responses rms SNRs for FFR at F_1

The results show major effects with the stimulus type ($F_{(1,11)}$=18.083, p<0.001) and stimulus content ($F_{(2,22)}$=20.242, p<0.001). However, the interaction between the type of stimulus and its content was not significant ($F_{(1.099,12.092)}$=1.224, p>0.1).

Next, the ANOVA was applied to the local rms-SNRs at F_1. The Mauchly's test revealed that sphericity was not present for the interaction between type of stimulus and its content (p>0.05). Therefore the confidence level p for the ANOVA was taken after the Greenhouse-Geisser correction. Again, we can see that there is a major effect of type of stimulus ($F_{(1,11)}$=19.002, p=0.001) and content ($F_{(2,22)}$=16.357, p<0.001) with the response at F_1 and there is no interaction between the type of stimulus and its content ($F_{(1.227,13.496)}$=1.994, p>0.1).

The following bar graphs (figures 4-21 and 4-22) summarize the results of the post-hoc pair-wise comparisons for amplitude and rms-SNRs of the FFR at F_1 across the stimulus variants for both quiet and noisy conditions. The detailed results of the statistical analysis for the response at F_1 can be found in Appendix F (Tables F-3, F-4, F-7, and F-8).

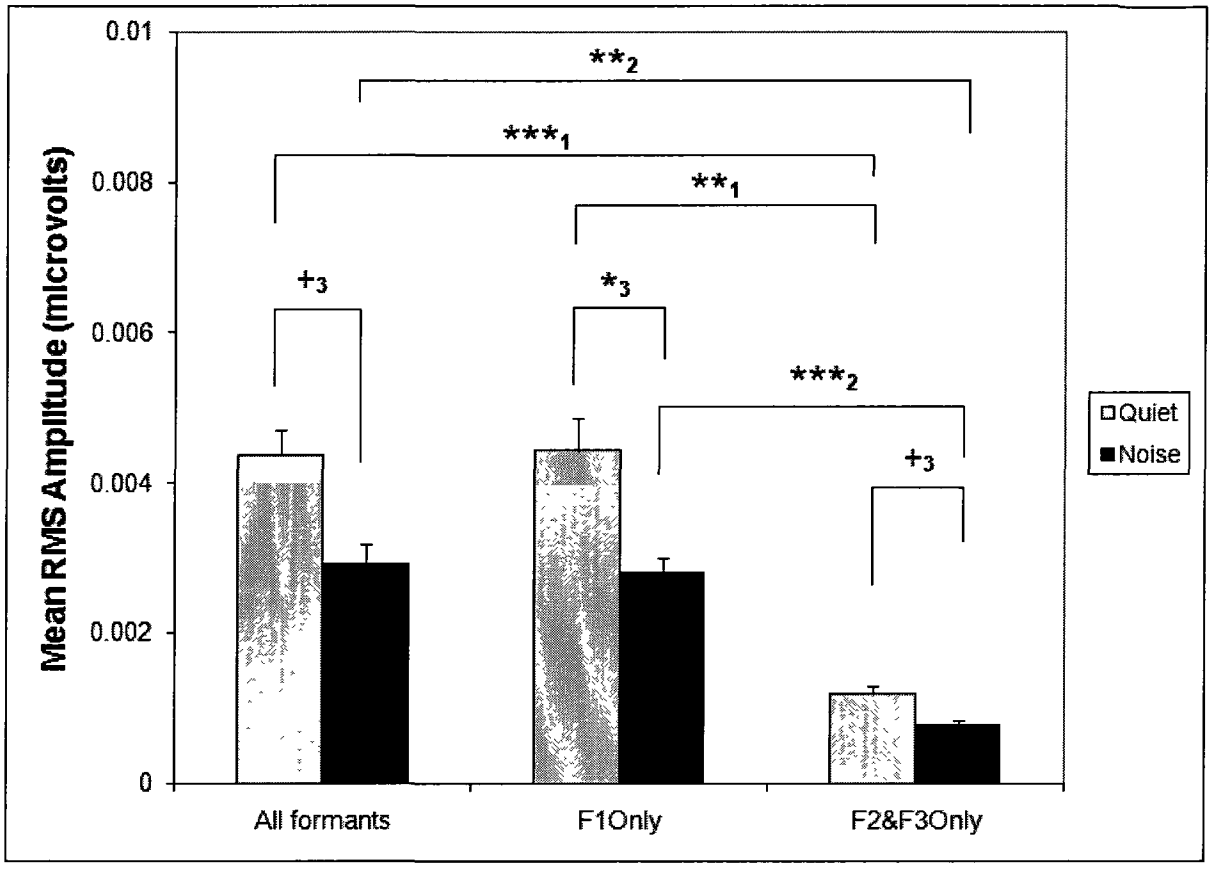

Figure 4-21: Comparison of the amplitudes of the responses at F_1 across the type of stimulus (quiet and noisy) and content (Allformants, F_1Only, and F_2&F_3Only)

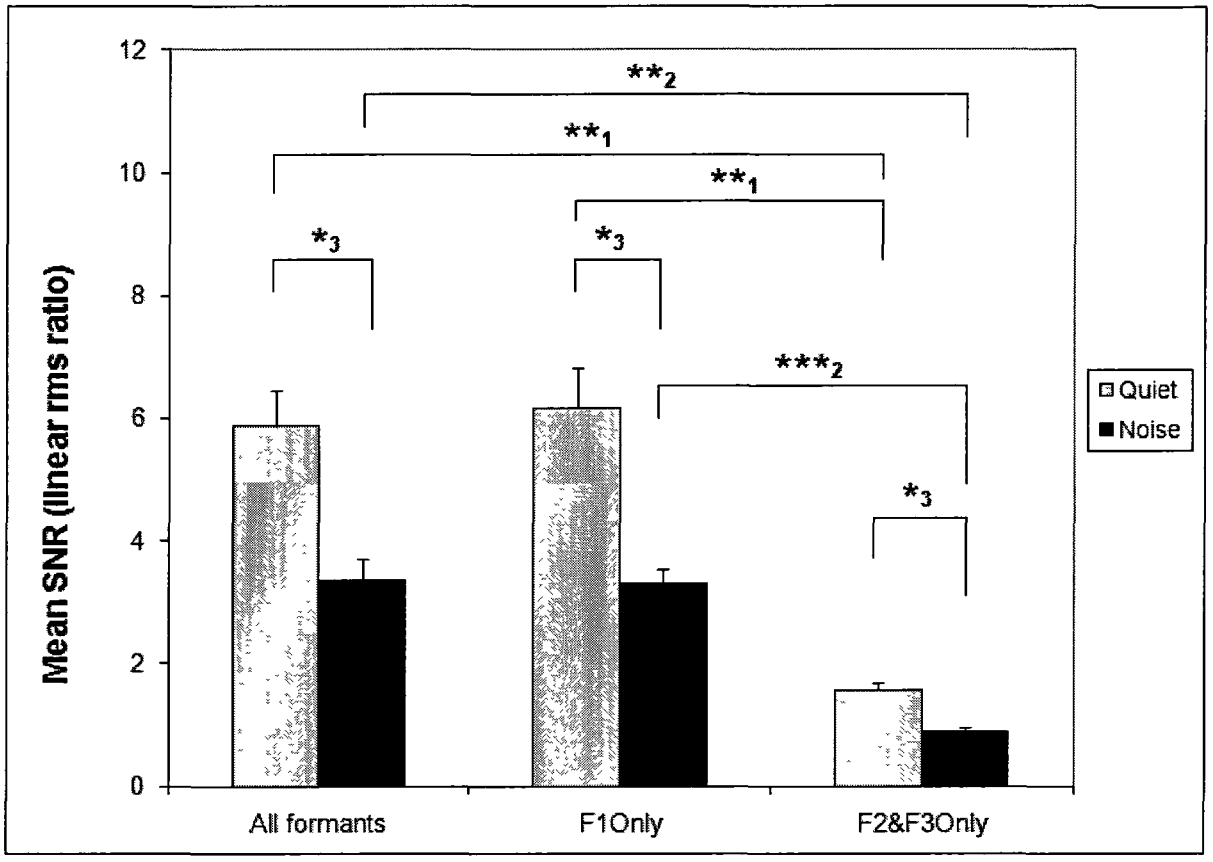

Figure 4-22: Comparison of the local rms-SNR of the responses at F_1 across the type of stimulus (quiet and noisy) and content (Allformants, F_1Only, and F_2&F_3Only)

The pair-wise comparisons showed that the amplitude of the FFR at F_1 generated using the clean F_2&F_3Only stimulus was weaker than the responses due to the Allformants and F_1Only stimuli ($p<0.01$). This result is expected because the component at F_1 in the F_2&F_3Only stimulus is weak. There was no difference between the means of the amplitude of the FFR at F_1 with the Allformants and F_1Only stimuli in either the quiet or noisy condition. When noise was present, the amplitude of the FFR at F_1 from the F_1Only stimulus was weaker than for the quiet condition ($p<0.05$). The amplitude of the response at F_1 using the Allformants noisy stimulus was reduced, although the reduction only tended towards statistical significance ($p=0.07$).

The pair-wise comparisons with the local rms-SNRs at F_1 gave similar results to the ones obtained with the amplitudes, although in this case the reduction of the rms-SNR obtained with the noisy Allformants stimulus was statistically significant ($p<0.05$).

4.4 Contribution of the Resolved and Unresolved Harmonic Regions to the Overall Response at F_0

We also looked at the grand-averaged vectors across subjects for all three stimulus variants (Allformants, F_1Only, and F_2&F_3Only). This was done by adding the vectors across the twelve subjects and dividing by the number of subjects. We also computed the Vector Sum of the F_1Only and F_2&F_3Only vectors for both the quiet and noisy conditions. Figures 4-23 and 4-24 show the results of this analysis. Qualitatively, in the quiet condition (figure 4-23), the vector of the response to Allformants is very different from the Vector Sum of the responses to the other two stimulus variants indicating that the contributions of the resolved and unresolved harmonic regions add up non-linearly when both are present in the stimulus. Table 4-1 shows the p-values resulting from the comparison between the magnitudes of the response to Allformants and the magnitude of the Vector Sum across subjects. The results (Table 4-1) show that the response at F_0 due to the Allformants and Vector Sum for the quiet condition are significantly different ($p<0.001$) again demonstrating that the responses at F_0 from the resolved and unresolved harmonic regions of speech do not add linearly when both regions are present in the stimulus. The results from the noisy condition (figure 4-24) show that the magnitudes of the Allformants and Vector Sum vectors are not statistically different ($p>0.05$). The magnitude of the F_1Only vector is similar to the Allformants vector suggesting that the response from the unresolved harmonics is not contributing significantly to the overall response at F_0 when noise is present.

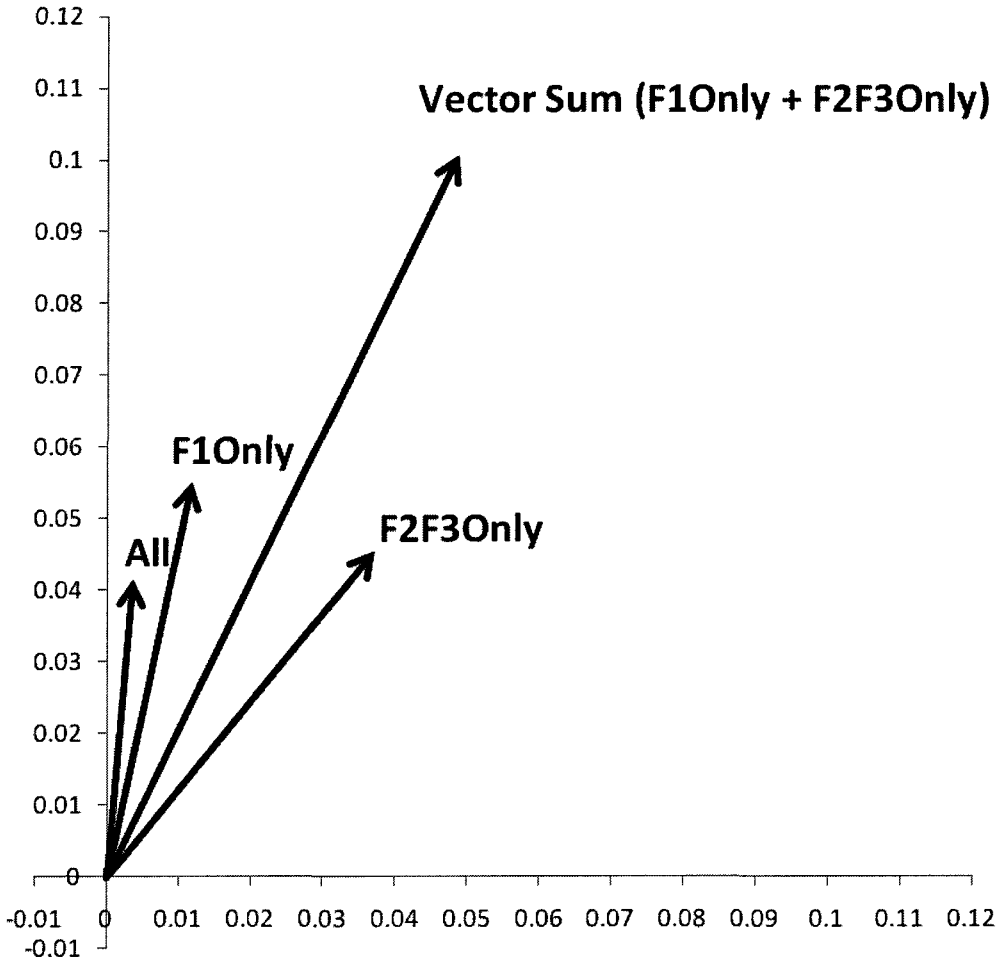

Figure 4-23: Vector Diagram of the EFR response at F_0 for the quiet condition

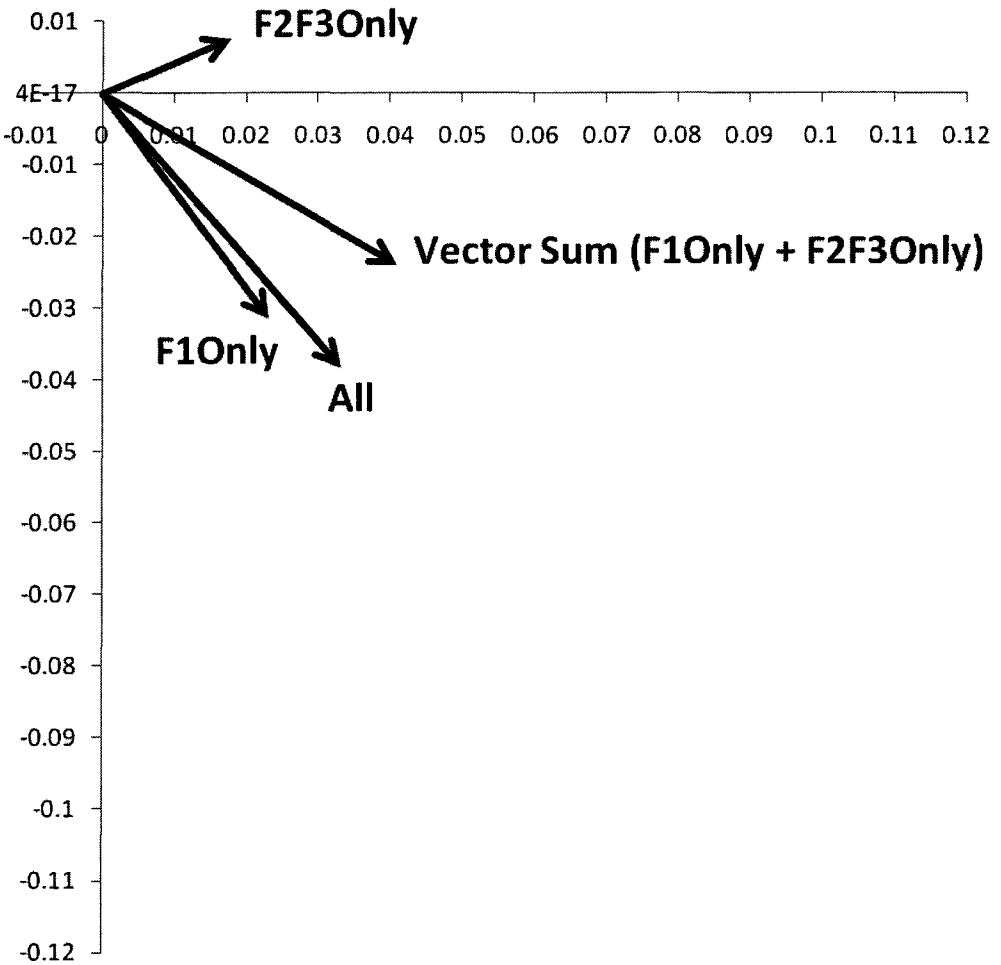

Figure 4-24: Vector Diagram of the EFR response at F_0 for the noisy condition

	p-value
Allformants_quiet vs Vector Sum_quiet	0.001
Allformants_noisy vs Vector Sum_noisy	0.532

Table 4- 5: Two-tailed t-tests results between vector magnitudes

4.5 F_0 Response Latency

In order to investigate the origin of the neural processes in the brainstem for the resolved and unresolved harmonics, the latencies between the acoustic stimulus and the evoked response at F_0 were estimated. This was done using the time-domain grand-averaged responses obtained using the three variants (Allformants, F_1Only and F_2&F_3Only). The results are reported in the following tables (including transmission delay of ~ 1ms):

	Quiet for F_0
Stimulus	Latency(ms)
Allformants	7.6458
F_1Only	7.7083
F_2&F_3Only	7.4375

	Noisy for F_0
Stimulus	Latency(ms)
Allformants	10.1458
F_1Only	10.3333
F_2&F_3Only	8.1250

Table 4- 6: Latency of the response at F_0 for the three stimulus variants in quiet and in noise

Chapter 5 : Discussion

5.1 Transient Response

The responses obtained showed that adding noise to a speech stimulus affects the transient response in terms of the latency and slope of the VA complex. With the addition of noise, the grand-averaged transient responses showed a steeper VA slope and a delayed VA onset, and the amplitude of the V wave also appears higher. The noise-induced delay of the VA onset (0.5-1 ms) agrees with what was found in previous studies (Russo et al., 2004; Skoe and Kraus, 2010). However, the steeper slope was not seen in previous studies, which in fact noted a reduction in the slope with the addition of noise. In this study, the steeper VA slope and higher amplitude of the peak V may have occurred because noise was not continuous between each presentation of the vowel (presentations were separated by approximately 33 ms of silence). Therefore, in our study the onset response could be due to a combination of the simultaneous onset of the underlying speech stimulus and of the added noise. In addition, the sudden onset of the noise could have resulted in a post-auricular artifact in the response, which represents the electrical activity evoked in the muscle located just behind the ear (the post-auricular muscle). Although having the reference electrode placed at the earlobe instead of the mastoid is expected to reduce the intensity of the artifact in the response, it is possible that this time-locked artifact was present as an onset response to noise and contributed to the steeper VA slope and higher amplitude of peak V with the noisy stimulus.

5.2 The Envelope Following Response (EFR) at F_0

5.2.1 Contribution of the Resolved and Unresolved Frequency Regions to the EFR at F_0 in Quiet

The magnitude of the EFR is based on various factors: the ability of the auditory neurons to phase-lock to the acoustic stimulus, how many neurons are firing at the fundamental frequency, and which regions in the cochlea actively respond to the stimulus frequency content (Greenberg et al., 1987).

The results obtained for the quiet condition showed that the EFR at F_0 generated solely from the resolved and the unresolved harmonics regions did not differ in terms of amplitude and

SNR. Obtaining responses due to both harmonic regions is not surprising given that the pitch is well represented in the temporal firing patterns of the peripheral auditory fibers across the full range of characteristic frequencies (see figure 2-7). However, based on psychophysical studies we would have expected that the response at F_0 from the unresolved harmonics region to be lower in strength, because the pitch of a harmonic complex is more salient perceptually for resolved harmonics than for unresolved harmonics (Shackleton and Carlyon, 1994; Cedolin and Delgutte, 2004). It is not clear why the strength of the evoked response does not follow perceptual saliency, but perhaps the perception of pitch depends on other forms of neural coding than the phase-locked activity that underlies the evoked response (e.g. rate-place coding) and likely involves many other cortical levels of speech processing (e.g. primary auditory cortex).

The grand-averaged responses obtained with F_1Only and F_2&F_3Only were also not different in magnitude from the one obtained with Allformants. This suggests that the responses due to the resolved and unresolved regions are combined non-linearly when they are both present in the stimulus.

To identify which region was contributing the most to the overall response at F_0, the vector analysis performed on the responses from F_1Only and F_2&F_3Only showed similar magnitudes, but not phase. Because of the difference in phase angles, we cannot conclude with certainty which contributes the most to the response at F_0 since further statistical tests must be done on both magnitude and phase simultaneously. The question of the origin of the responses at F_0 was instead investigated based on latency as described below.

5.2.2 Effect of Noise on the EFR at F_0

The EFR responses at F_0 obtained from the Allformants stimulus were not statistically different between the quiet and noisy conditions. This robustness to noise is consistent with the results of previous studies (Cunningham et al., 2001; Cunningham et al., 2002; Russo et al., 2004; Johnson et al., 2005; Parbery-Clark et al., 2009), and is probably explained by the contribution of all regions of the stimulus spectrum to the response. Even at the peripheral level, the fundamental periodicity is well represented in the interspike intervals recorded

from the auditory nerve fibers with both low and high characteristic frequencies (Cariani and Delgutte, 1996).

The responses generated from the resolved vs. unresolved harmonic regions of the stimulus depicted a more complex picture. The response to the higher frequency unresolved harmonics region (i.e. with the F_2&F_3Only stimuli) was significantly affected by noise which is supported by perceptual studies (Assmann and Summerfield, 2004). On the other hand, the response at F_0 to the resolved harmonics regions remained robust to the addition of noise. The susceptibility to noise in the response from the unresolved harmonics at F_0 versus the one from the resolved harmonics may be explained from the difference in the effective SNR level in each variant.

However if the response to the resolved harmonics is intrinsically more robust than to the unresolved, this may start right at the peripheral level, because when neural activity was recorded in auditory nerve fibers, low frequency harmonics were found to generate interspike intervals at the fundamental periodicity with stronger peak-to-background ratios than higher frequency harmonics (Cariani and Delgutte, 1996). A difference in susceptibility to noise agrees with previous results obtained in perceptual experiments on the discrimination of harmonic complexes in the presence of overlapping maskers (Carlyon, 1996). Resistance to noise is probably explained by a strong phase-locking property of the auditory-nerve fibers for the duration of the stimulus, and which was not affected by the level of noise added in this study. If the SNR in the stimulus were made significantly lower, we would expect their neural phase locking to degrade to the point where the measured evoked response due to the resolved harmonics becomes weakened.

The results of the vector analysis gave similar conclusions. When noise was added, the magnitudes of the Allformants and Vector Sum vectors are not significantly different. This is probably because the response to F_2&F_3Only is significantly weaker in noise, making the response due to F_1Only dominant when both regions are present in the stimulus.

There is an interesting question as to whether the auditory system is performing noise reduction when the stimuli are noisy. One way to address this question is to measure how the response amplitudes and SNR vary with changing stimulus SNR. Because in this study only one noise level was tested, this question is difficult to answer now, and instead will be the subject of future work.

5.2.3 Origin of the EFR Response at F_0

The latencies obtained between the stimuli and the responses are similar for all the stimulus variants for the quiet condition. This suggests that the content is encoded in the same or closely located area in the brainstem. The latency results for the quiet condition are similar to those obtained in previous studies which estimated an average latency of 7.7 ms with the synthetic consonant-vowel /da/ (Parbery-Clark et al., 2009; Skoe and Kraus, 2010). This suggests an origin for the response at F_0 in the upper brainstem (and in particular the inferior colliculus), which has been proposed as the main generator for speech evoked ABRs (Johnson et al., 2008). For the noisy condition, the latency increased which is expected because noise will cause a delay in the response (Parbery-Clark et al., 2009). The latency from the F_2&F_3Only stimulus in noise is significantly shorter, but the response at F_0 produced from this stimulus is quite weak in comparison to the responses to the other stimuli and relative to the surrounding noise, so the obtained latency may not be reliable.

The vector analysis of the responses at F_0 depicts a more complicated picture about the origin of the response. The response to Allformants (which contains both resolved and unresolved harmonics) in quiet is much smaller than the vector sum of the responses to F_1Only and F_2&F_3Only, despite the Allformants stimulus being equal to the sum of the F_1Only and F_2&F_3Only stimuli. This suggests that the responses at F_0 due to the resolved and unresolved harmonics interact non-linearly at some point before reaching the upper brainstem or within it. One possibility that we considered was that the non-linearity was due to saturation in the neural activity that limits the maximum possible amplitude of the response with the Allformants stimulus. However, this explanation is unlikely given that other studies have reported responses with higher amplitudes (above 0.1μV) (Parbery-Clark et al., 2009).

The similarity between the response vectors to Allformants and F_1Only stimuli (see figures 4-23 and 4-24), may indicate a possible common physical origin for these two responses. The vector diagrams therefore suggest that when the unresolved harmonics (corresponding to the F_2&F_3Only variant) are present in the stimulus, there is a non-linear interaction with the response to the resolved harmonics (corresponding to the F_1Only variant), in which the latter response dominates.

The vector analysis also shows a difference in phase between the F_1Only and F_2&F_3Only vectors. This difference might be caused by different time delays in the processing of the resolved and unresolved harmonics, possibly starting at the level of the cochlea.

The pitch of resolved and unresolved harmonics: One mechanism or two?

It is unclear if the pitch of resolved and unresolved harmonics is processed using a single or two separate mechanisms in the auditory system. A number of previous psychophysical studies suggest that there are two independent mechanisms (Carlyon and Shackleton, 1994; White and Plack, 1998; Grimault et al., 2002). For example, White and Plack (1994) found that the integration time for resolved harmonics is shorter than for unresolved harmonics, and increasing the duration of the stimulus improves the accuracy for the pitch for unresolved harmonics. In contrast, other psychophysical studies have questioned the existence of two processing mechanisms for resolved and unresolved harmonics (Gockel and Carlyon, 2004; Micheyl and Oxenham, 2004). These studies showed that pitch discrimination does not differ between resolved and unresolved harmonics indicating that pitch is processed the same way for both harmonic regions. Moreover, a study using Functional Magnetic Resonance Imaging (fMRI) did not find a correlation between the activation level of relays in the auditory pathway up to the primary auditory cortex and the divergent perceptual pitch salience due to resolved and unresolved harmonics (Penagos et al., 2004). This study therefore suggests that a single subcortical mechanism from the cochlear nuclear nucleus to the inferior colliculus might be involved in the extraction of pitch for resolved and unresolved harmonics (Penagos et al., 2004).

The analysis performed on the responses collected in this study suggests the existence of two pathways for pitch extraction from resolved and unresolved harmonics, after which they converge in one area or two closely related areas. Starting in the periphery, it is well established that resolved and unresolved harmonics are processed in different regions of the basilar membrane and so are encoded in different groups of auditory nerve fibers. On the other hand, the near identity of our estimates of the response latencies due to the resolved and unresolved harmonics suggests that the recorded responses originate from a region of convergence in the upper brainstem. Therefore, two separate pathways are possible between the periphery and the area of convergence in the upper brainstem, but it is likely that there are non-linear interactions between them as suggested by the vector analysis of the responses.

Separate mechanisms for pitch processing from resolved and unresolved harmonics might not be entirely restricted to the peripheral and the brainstem level. With low and high-pitch missing-fundamental tones, Winkler et al. showed that there are two separate processes for resolved and unresolved harmonics in the auditory cortex (Winkler et al., 1997). Their investigation focused on how the duration of the stimulus affected pitch discrimination at the cortical level. They concluded that pitch could be extracted and discriminated from the low-order resolved harmonics by pattern-recognition processes in the cortex. Note that other potential sites located in the left hemisphere of the cortex might be involved in the neural processing of resolved and unresolved harmonics as well (Hall and Plack, 2008).

To summarize the ideas discussed above, figure 5-1 depicts a plausible conceptual model of the origin of the response to the pitch of resolved and unresolved harmonics. It suggests that the pitch of resolved and unresolved harmonics is processed separately, through different pathways beyond the periphery. Convergence of processing then takes place and is present in the upper brainstem. Beyond the periphery, there is non-linear interaction between the processing of resolved and unresolved harmonics, but the precise locus of these interactions is unknown.

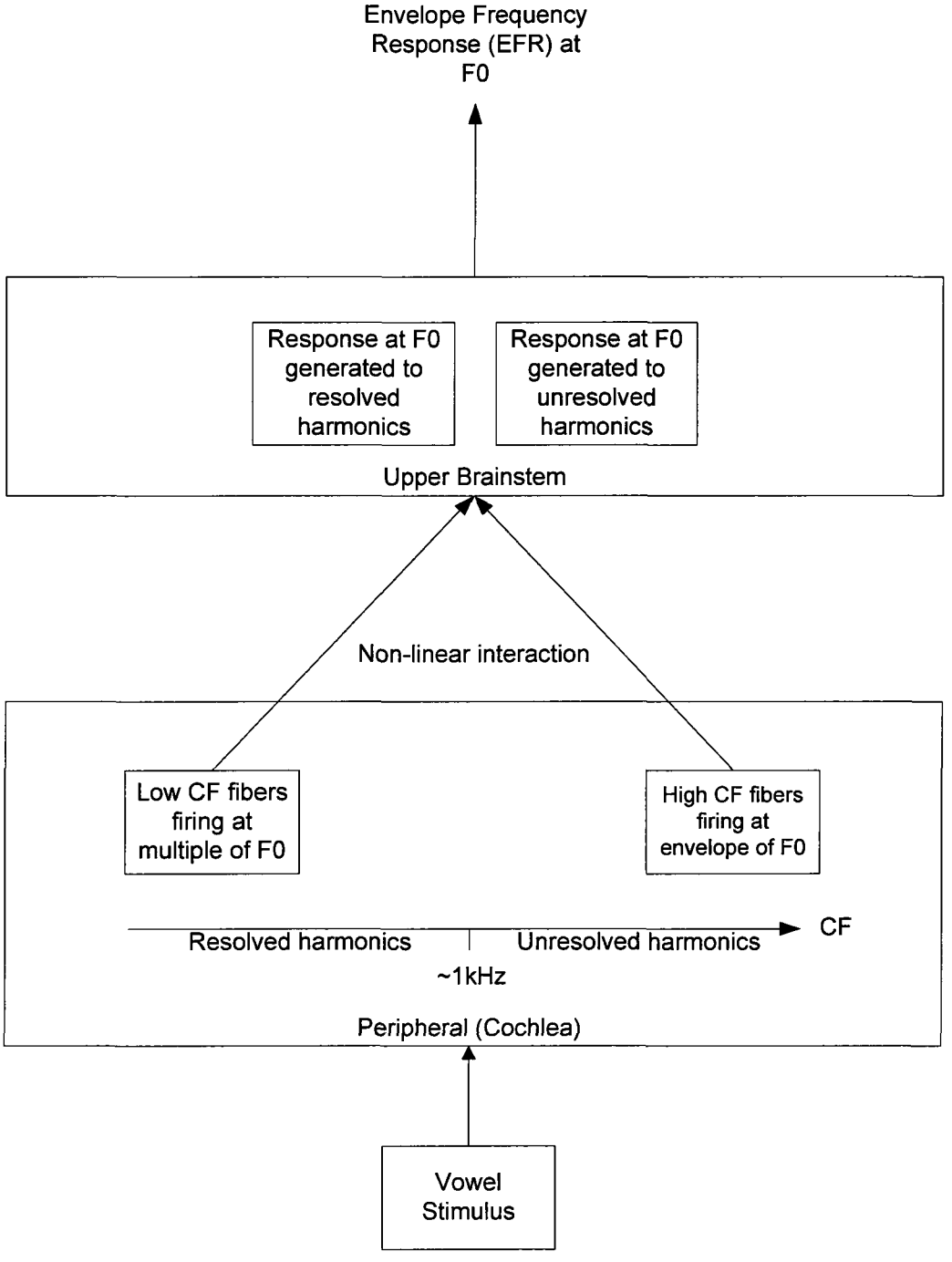

Figure 5-1: Conceptual model suggesting the origin of the response to the pitch of the resolved/unresolved harmonics in the auditory system

5.3 The Frequency Following Response (FFR) at F_1

5.3.1 Contribution of the Different Frequency Regions to the FFR at F_1 in Quiet

The results showed that the FFR at F_1 was weaker for the F_2&F_3Only stimulus as opposed to the Allformants and F_1Only stimuli for the quiet condition. This was expected because the component at F_1 (700 Hz) for the F_2&F_3Only stimulus is weak, as can be seen in the frequency spectrum of F_2&F_3Only (figure 3-7). Both Allformants and F_1Only stimuli contain a strong F_1 component which produces a well-defined FFR at F_1.

5.3.2 Effect of Noise on the FFR at F_1

The results of this study show that the FFR at F_1 is more susceptible to noise than the EFR at F_0, as reflected by the lower amplitudes and SNRs obtained for the noisy condition. This is consistent with the findings of previous studies (Cunningham et al., 2002; Russo et al., 2004). The statistical analysis performed on the grand-averaged responses showed that the response from the F_2&F_3Only stimulus was not statistically different from the surrounding noise.

The basis behind the generation of EFR and FFR may help explain the robustness of the EFR at F_0. It is the entire frequency range of the stimulus components that contributes to the EFR at F_0, while the FFR at F_1 is due to a specific single frequency component.

Chapter 6 : Conclusions and Future Work

6.1 Outcomes

A primary objective of the present research was to study the robustness of the speech brainstem response to noise. This was done by investigating the transient response, and the responses at the fundamental frequency (F_0) and the first formant (F_1) to the synthesized vowel /a/ in both quiet and noisy conditions. Another objective of the research was to study the contribution of the resolved and unresolved harmonics to the response at the fundamental frequency.

The analysis performed on the transient response for the quiet and noisy conditions showed an increase in the slope of the VA complex with the addition of noise. In previous studies, the slope was found to decrease in the presence of noise. We speculate that having intermittent noise in our study, versus continuous noise in the earlier work, explains our results.

For the steady-state response, the results show that in quiet, the resolved harmonics (F_1Only stimulus) and the unresolved harmonics (F_2&F_3Only stimulus) result in responses at F_0 that are equally strong, and similar to the response that results from the stimulus containing both resolved and unresolved harmonics (Allformants). In noise, the response at F_0 due to the resolved harmonics is much more robust than the response due to the unresolved harmonics, although the difference in robustness could be due to the difference in effective SNR between each of the noisy stimulus variants. In addition, the latency analysis suggests that the response at F_0 to both resolved and unresolved harmonics originates from the same region in the brainstem, although there appears to be non-linear interactions between the two responses. As for the steady-state response at F_1, it was reduced in noise for the Allformants and F_1Only conditions, unlike the response at F_0.

This work also proposed a plausible conceptual model to describe the origin of the responses at F_0 due to resolved/unresolved harmonics in the auditory system. The model suggests two separate pathways for processing of the pitch of resolved and unresolved harmonics that

converge in the upper brainstem, with non-linear interactions present somewhere along the two pathways.

6.2 Thesis Contributions

The following list summarizes the main contributions of this thesis:

- This work is the first to study the contributions of resolved and unresolved harmonic regions in speech to the speech auditory brainstem response in quiet and in noise. The findings were summarized in a conceptual model of the origin of the response at F_0 due to resolved and unresolved harmonics. The results of this study could contribute to a better understanding of pitch processing in the auditory system, a topic of much controversy. Moreover, an improved comprehension of the mechanisms involved in the generation of speech ABR would help in future clinical applications.
- The study confirmed previously published results by one research group that showed a higher robustness of the response at F_0 compared to the response at the first formant (F_1). This corroboration is important because the differential robustness of these two responses has been suggested as a possible marker for central processing disorders in children.
- The study involved the implementation of a vowel formant synthesizer, and experimental techniques that will be useful in future research projects.

6.3 Future Work

Since this study was based on one synthesized vowel, we cannot be certain that the results obtained are universally true for all types of speech sounds. This work was also based on stationary white Gaussian noise, and comparison with the effects of other types of noise such as shaped noise to study SNR effects in different frequency regions and multi-speaker babble noise could be useful. Future work therefore includes:

6.3.1 Continuous vs. Interrupted Noise

The study the effect of continuous vs. interrupted noise on the transient response (specifically on the VA complex) may clarify the cause of the increase of the VA slope

found in this study. Previous studies had found a reduction of the slope in the presence of noise in the stimulus (Russo et al., 2004; Skoe and Kraus, 2010) .

6.3.2 Comparisons with Other Synthetic Vowels

Further investigations using other synthetic vowels as stimulus should be performed, to see if comparable results are obtained. This could possibly complement the model of the response at F_0 due to resolved/unresolved harmonics.

6.3.3 Investigations with Modulated Noise

Rapid changes in sound amplitude occur at different time scales in the speech signals, and so are of interest when studying speech. Some previous studies have demonstrated that adding continuous random noise to a weak periodic signal stimulus aided its detection and neural representation in the auditory response. This enhancement is believed to result from adaptive mechanisms in the auditory system such as peripheral operating range shifts of the auditory nerve fibers (Frisina et al., 1994; Frisina et al., 1996; Cunningham et al., 2002). In a natural acoustic environment, the noise amplitude and overall spectral pattern usually varies with time. Preliminary work has been done on the generation of the vowel stimulus with modulated noise added. However, further investigations indicated that pursuing the modulated noise design would add considerable complexity to the interpretation of the results (including issues related to temporal resolution, forward and backward masking, etc.). As a result, the experimental design in this thesis focused on stationary noise, and the study of modulated noise was left for future work.

Appendix A: Calibration using the Sound Level Meter 2235 Brüel & Kjær

First, the SLM is calibrated (measuring in slow-mode) and operating in the 60-120 range. Calibrating the SLM is done using the pistonphone placed on the microphone. A screwdriver provided with the instrument should be used to adjust the SLM to 124.0 dB SPL using the first of the three holes on the right side of the SLM case. After calibrating the SLM, it is set up as shown below to calibrate the sound generated by the Bio-logic® instrument. The measurements are done using the Bio-logic® insert earphone's tube inserted in the 2cc coupler placed on top of the 4144 microphone and onto the artificial ear connector (type 4152). Then the SLM is turned on and the sound level is measured.

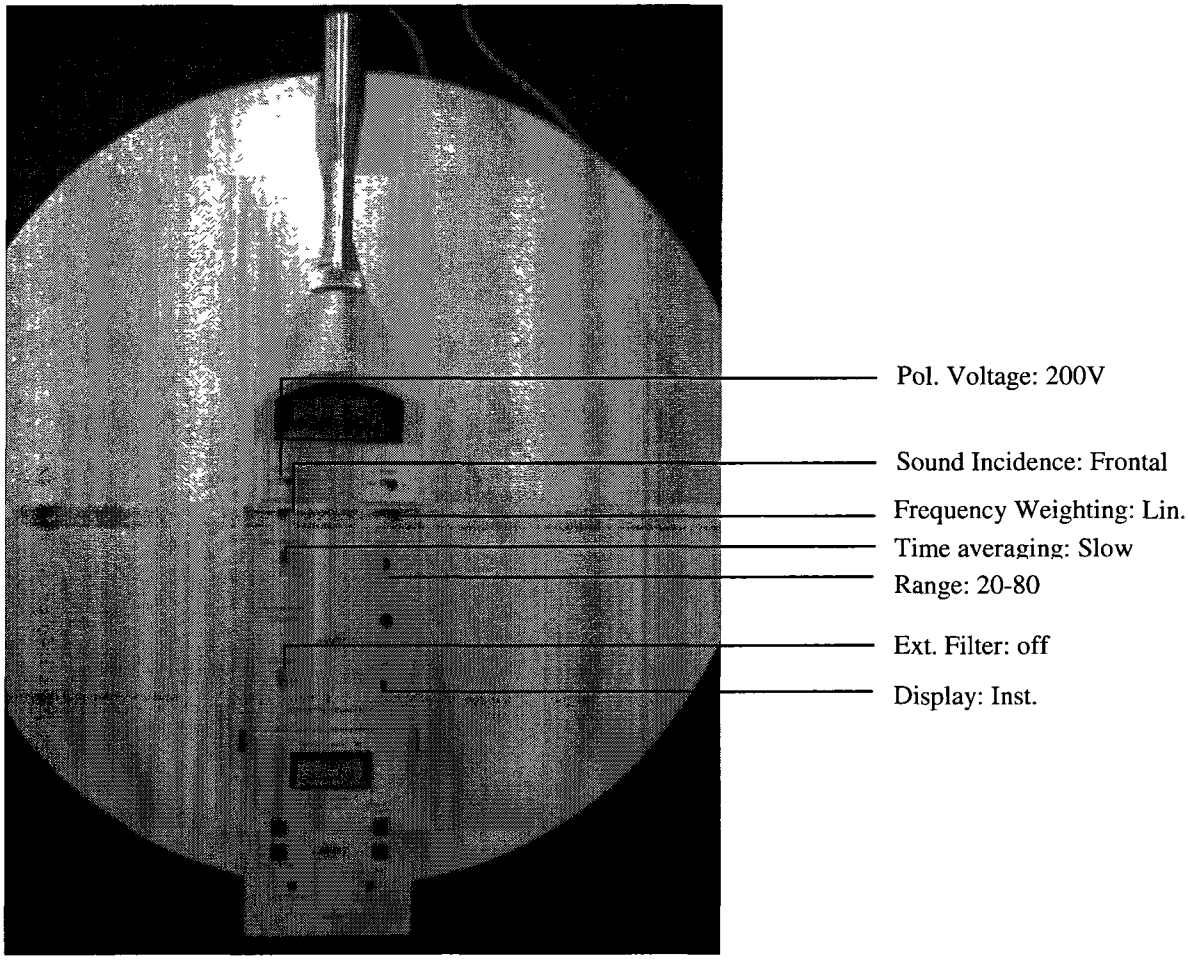

Figure A- 1: Calibration of the generated sound using the Sound Level Meter

Below are the results obtained from the calibration of the system to the noisiest stimulus. The noisiest stimulus was the 300ms-vowel /a/ with stationary white Gaussian noise with SNR of -5dB (all formants included). The internal amplification factor in the BioMARK™ system was set to give a maximum intensity level of 79.5 db SPL for this stimulus. This setting remained unchanged when presenting the remaining stimuli.

Stimulus	BioMARK™ (dB)	SLM (dB SPL)
Stationary noise, SNR: -5 dB	98*	79.5
Stationary noise, SNR: -5 dB F1		78.7
Stationary noise, SNR: -5 dB F2&F3		76.2
Stationary noise, SNR: 10000 dB		76.7
Stationary noise, SNR: 10000 dB F1		75.6
Stationary noise, SNR: 10000 dB F2&F3		67.4

Table A- 1: Calibration results

*This is not the level of the output signal at the earphone, but this is an internal amplification factor for BioMARK™.

Appendix B: Tests for Electromagnetic Leakage, and Sample Tests for Power-line Noise Contamination

Testing for electromagnetic leakage was performed before starting the evoked response recordings. This test is used to verify that the BioMARK™ system does not introduce undesired electric potentials that leak from the stimulus signal in the sound generating equipment to the recorded response. This procedure was completed using a Bio-logic® insert earphone inserted into a G.R.A.S. coupler which was connected to the turned-off SLM, type RA0045 (according to the IEC711 standard), and electrodes positioned on the subject's scalp. Apart from connecting the earphone to the coupler, the measurement proceeded in exactly the same way as when running a "real" measurement of the evoked response. The stimulus used for testing leakage was the 300ms clean /a/ vowel and the responses were averaged over 3000 realizations. The evoked response obtained is shown in figures B-1 (time domain) and B-2 (frequency domain).

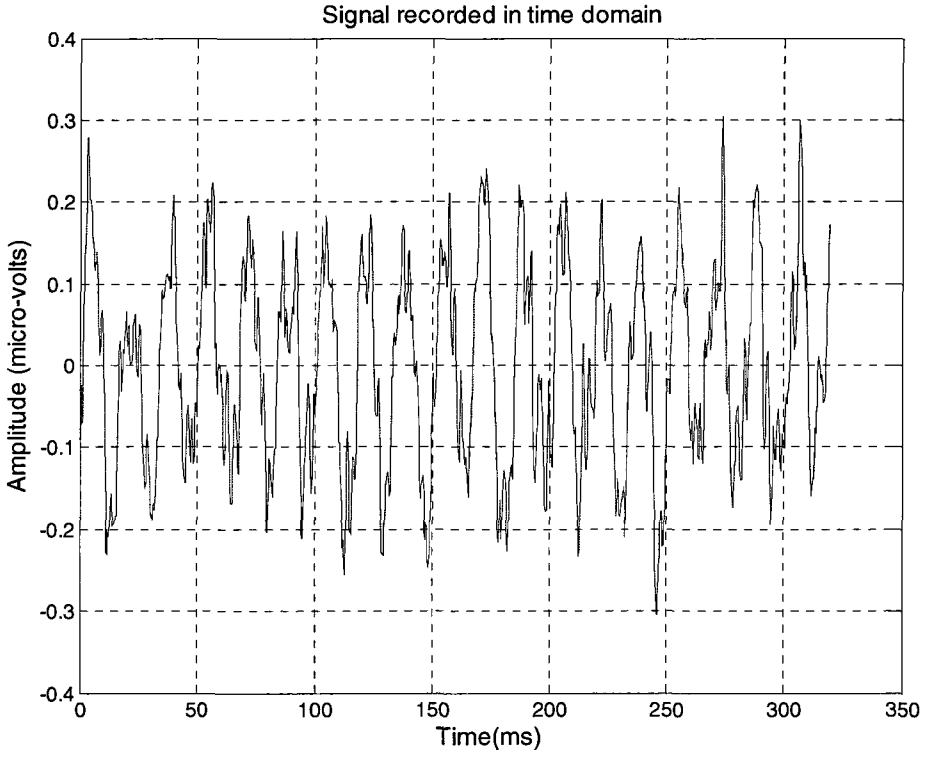

Figure B- 1: Test for electromagnetic leakage-Connection with coupler (time domain)

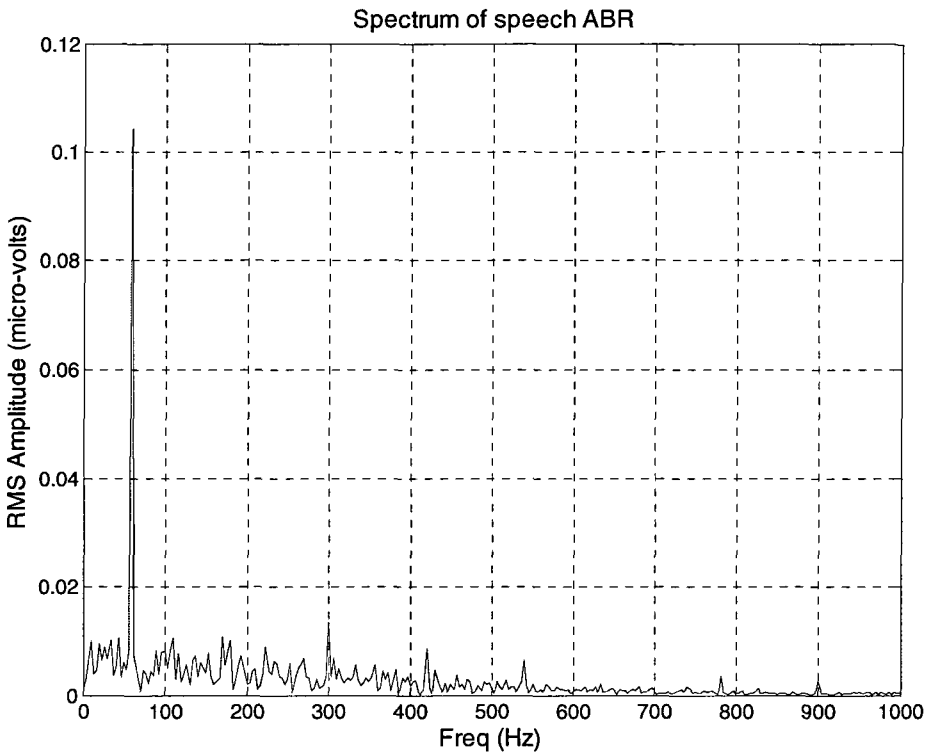

Figure B- 2: Test for electromagnetic leakage-Connection with coupler (frequency domain)

The frequency spectrum of the evoked response showed a prominent peak at 60Hz, which was due to noisy equipment present in the audiology booth at that time. (A deeper investigation is presented in section 3.7 of this report.) The measurement with the coupler demonstrates that there is no evidence for electromagnetic leakage of the stimulus into the recorded response, because there are no components in the spectrum at F_0 or its harmonics (excluding frequencies that coincide with the harmonics of 60 Hz). For comparison, the evoked response obtained when the subject was listening to the stimulus is shown in figures B-3 (time domain) and B-4 (frequency domain).

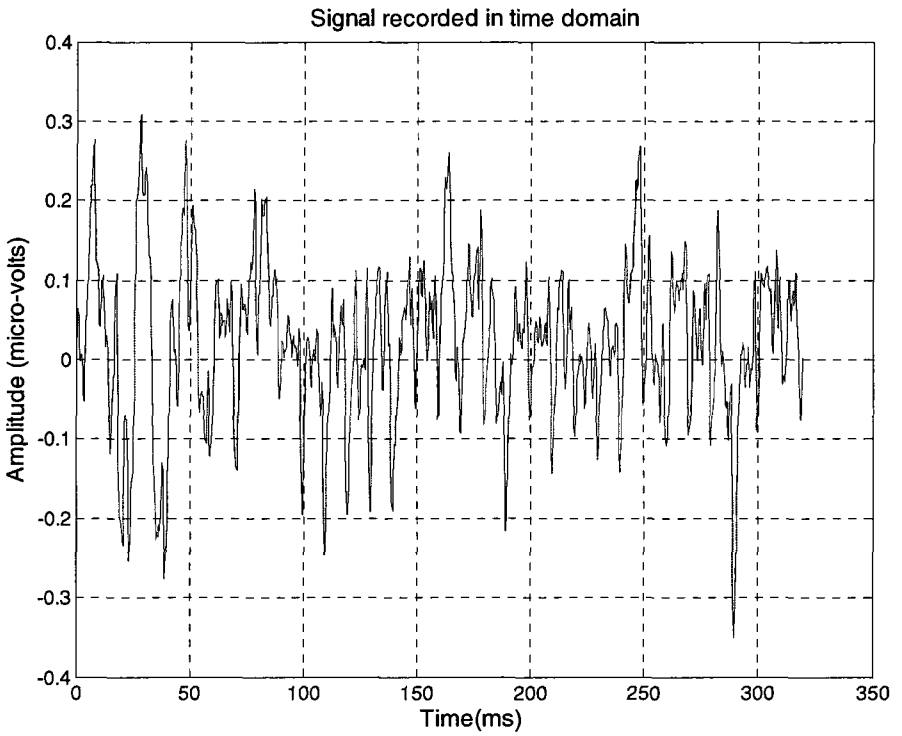

Figure B- 3: Test for electromagnetic leakage-Connection with subject (time domain)

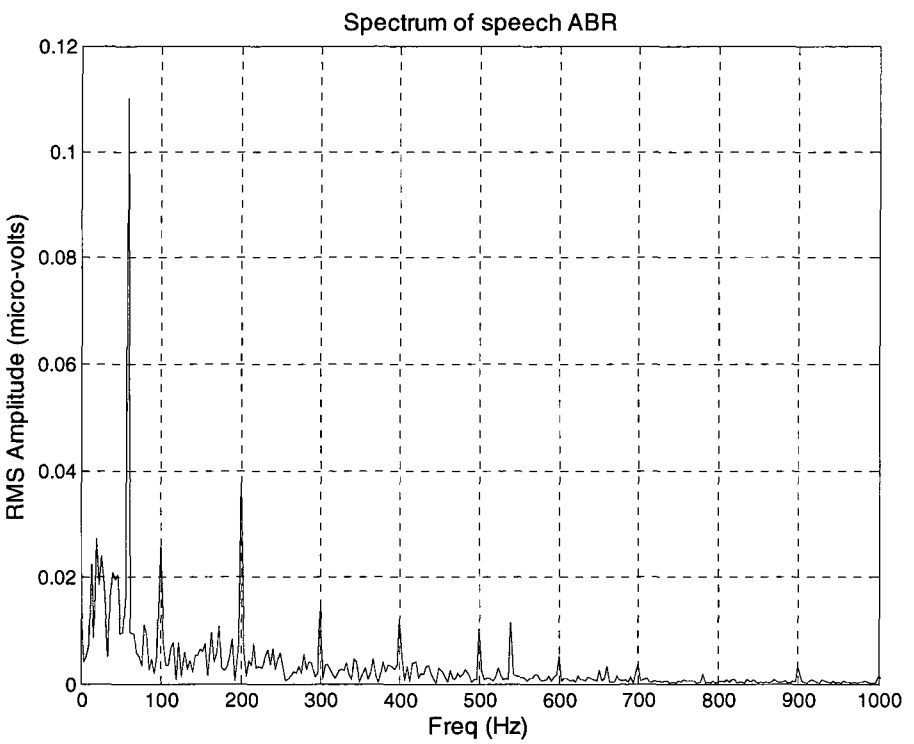

Figure B- 4: Test for electromagnetic leakage-Connection with subject (frequency domain)

Test for power-line contamination

Numerous tests were done to investigate the sources of the power-line contamination (at 60 and its harmonics); here are two of them.

Outside the audiology booth, a PC was used to run the movie on a flat screen mounted on the wall inside the booth. We suspected that the power cable to the screen was emitting a powerful 60 Hz signal that was contaminating the responses. Tests were made using a clean stimulus (SNR10000) when a movie was running on the booth's screen versus a movie playing on a laptop.

The spectrum of an evoked response with a muted movie running on booth's screen is shown in figures B-5 and B-6:

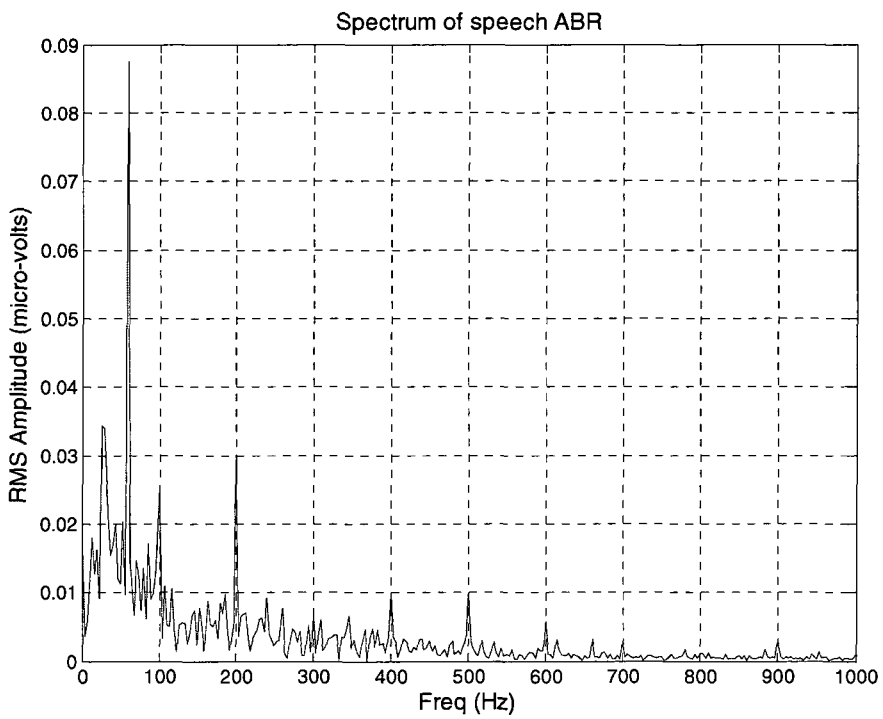

Figure B- 5: Test for power-line contamination-1st attempt

F_0:

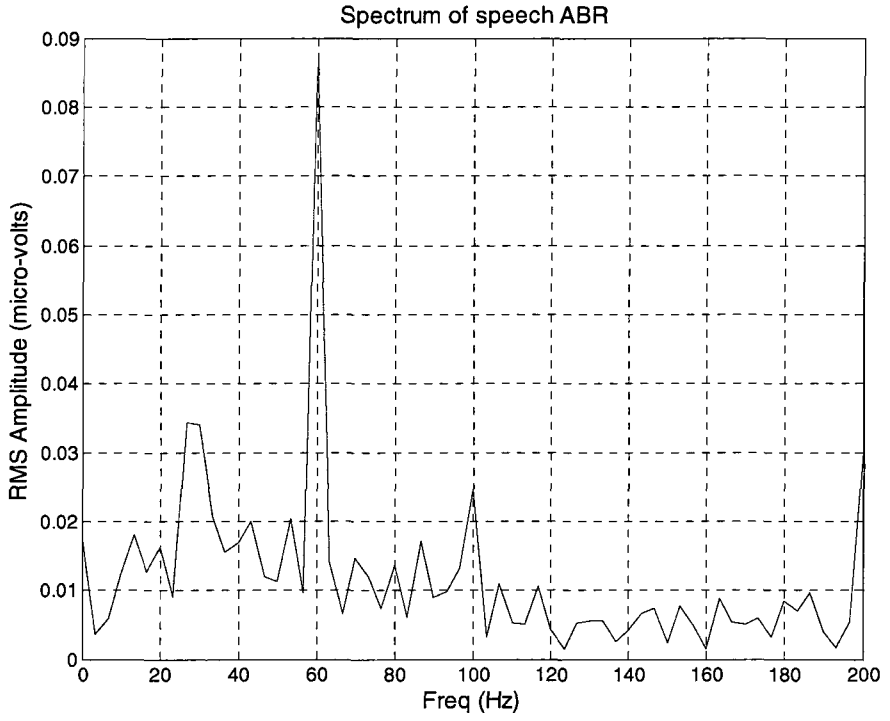

Figure B- 6: Test for power-line contamination-response in the region of F_0 (zoomed)

Power-line noise contamination is present in the spectrum, as shown in figure B-6 by the significant peak at 60 Hz. The spike is stronger than the response at F_0 (100 Hz).

The VA complex of the transient response in the time domain is shown in figure B-7:

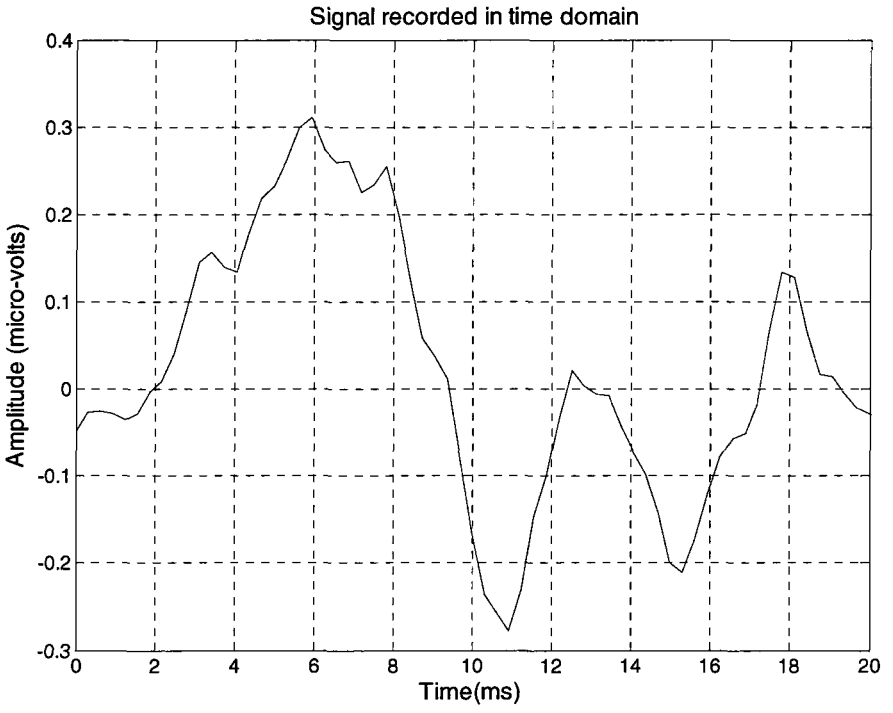

Figure B- 7: Test for power-line contamination on the transient response (zoomed)

Since power-line noise contaminated the recordings in this case, it likely also affected the transient response.

The second test was aimed at reducing the effect of the power-line noise on the recordings by running the movie on a laptop connected to the iso-transformer without the booth's screen.

Figures B-8 and B-9 show the spectrum of the response obtained with the movie running on the laptop, with the booth's screen unplugged and outside PC turned off:

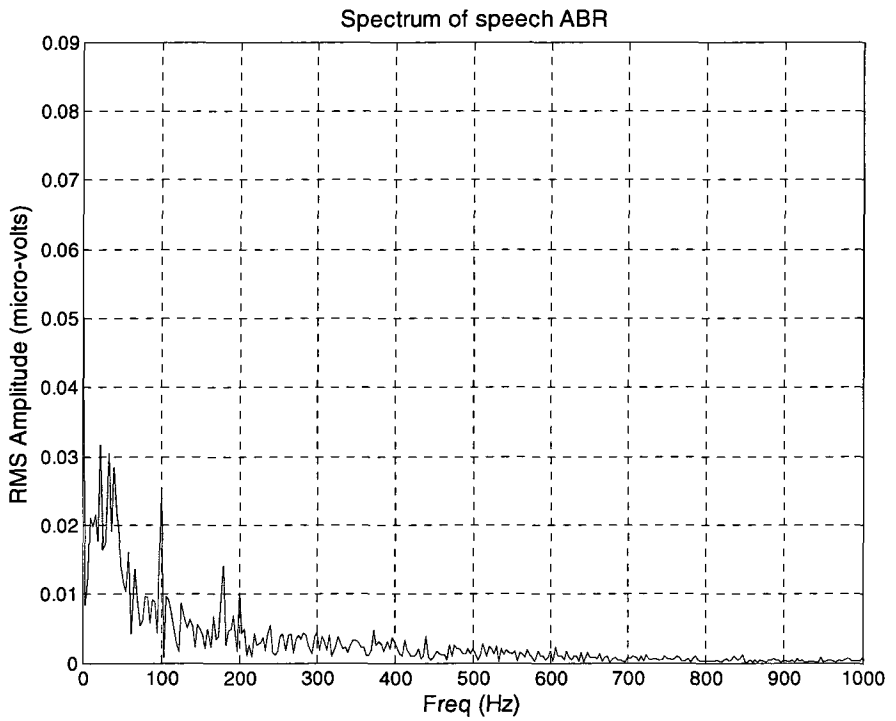

Figure B- 8: Test for power-line contamination-2nd attempt

F_0:

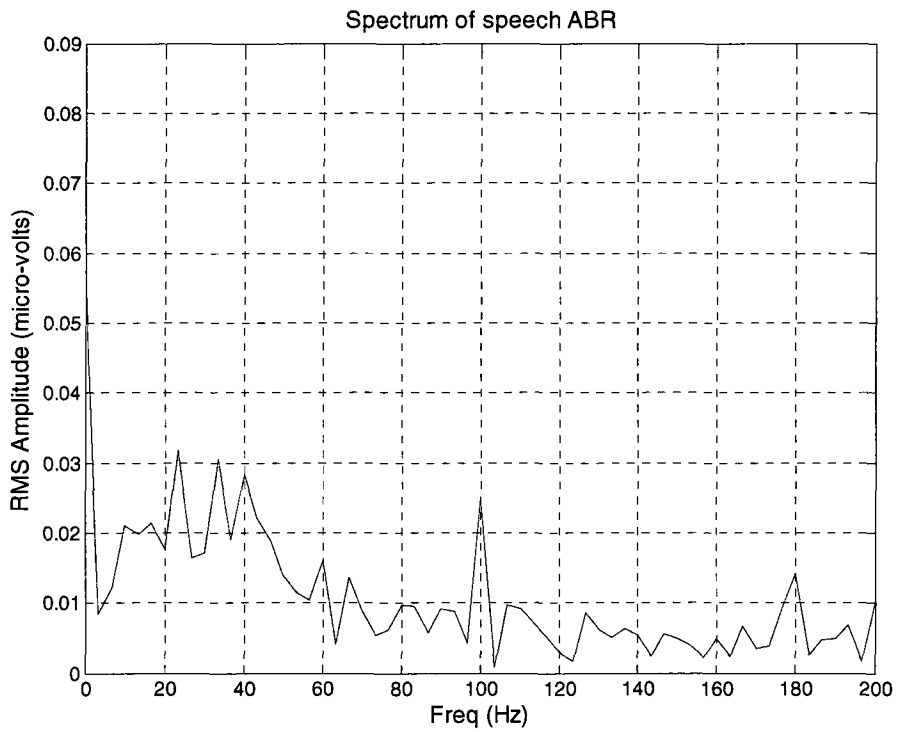

Figure B- 9: Test for power-line contamination-response in the region of F_0 (2nd attempt, zoomed)

Figure B-9 shows that the method used for this test was successful in reducing the contamination of the power-line noise.

The VA complex of the transient response in the time domain is shown in figure B-10:

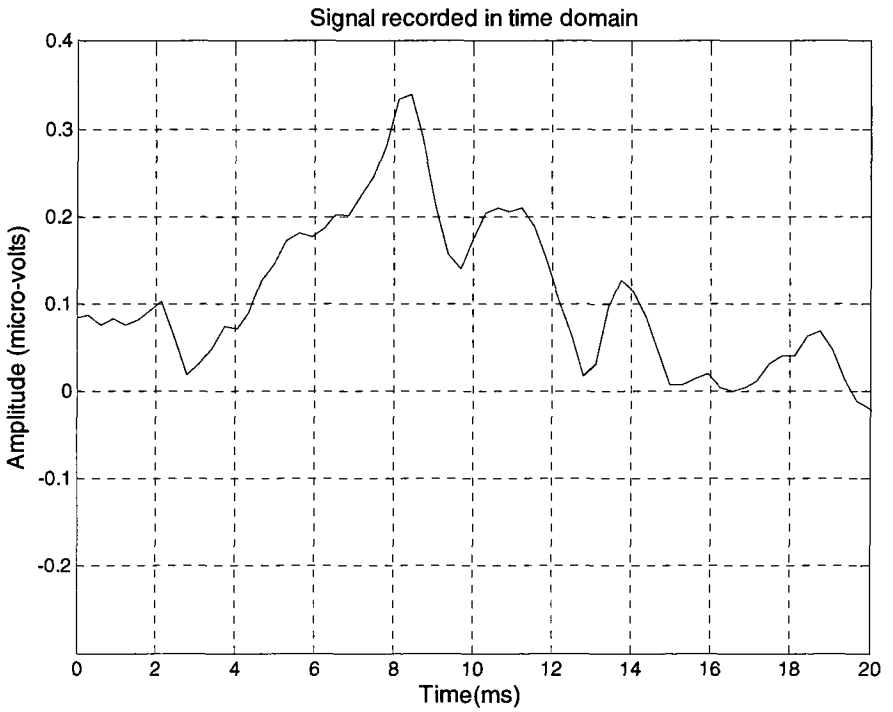

Figure B- 10: Test for power-line contamination on the transient response (2nd attempt, zoomed)

The VA complex is better defined in figure B-10 around 10 ms as compared to the recordings of the previous test. It also compares well with the transient response reported in the literature. Performing these two tests along with several others helped reduce power-line noise contamination. Although we were not always successful in suppressing this contamination, our measurements of the steady-state response amplitudes were unaffected since the F_0 and F_1 frequencies (100 and 700 Hz respectively) are not harmonics of 60 Hz. Moreover, the harmonics of 60 Hz were also excluded in the estimates of the local SNR.

Appendix C: Amplitude and SNR of the Responses at F_0

Amplitudes:

	Quiet for F0		
Subjects	Allformants	F1Only	F2&F3Only
1	0.007096	0.044228	0.055234
2	0.068968	0.116311	0.105377
3	0.101358	0.077351	0.124194
4	0.048684	0.059390	0.045877
5	0.031574	0.036332	0.046387
6	0.081196	0.037371	0.026783
7	0.081534	0.084391	0.135919
8	0.052903	0.066492	0.047007
9	0.014011	0.008886	0.021545
10	0.052307	0.092408	0.062732
11	0.037820	0.063950	0.052931
12	0.130035	0.143111	0.095688
rms mean (micro-volts)	0.05896	0.06919	0.06831
standard deviation	0.03574	0.03693	0.03767
standard error of mean	0.01032	0.01066	0.01087
Grand-average of signals (rms mean micro-volts)	0.04173	0.05644	0.05860

	Noisy for F0		
Subjects	Allformants	F1Only	F2&F3Only
1	0.062490	0.065525	0.031114
2	0.089336	0.075577	0.032282
3	0.027239	0.013467	0.004077
4	0.004287	0.023079	0.014270
5	0.073339	0.053591	0.040953
6	0.052159	0.013947	0.012284
7	0.139898	0.122962	0.058723
8	0.043414	0.044279	0.017270
9	0.047248	0.049998	0.029856
10	0.078425	0.058898	0.029201
11	0.079017	0.082945	0.031009
12	0.045886	0.056282	0.025137
	0.06189	0.05505	0.02718
	0.03433	0.03093	0.01439
	0.00991	0.00893	0.00415
	0.05040	0.03888	0.01917

Table C- 1: Amplitude of the response at F_0 (all subjects)

SNRs:

	Quiet for F0		
Subjects	Allformants	F1Only	F2&F3Only
1	0.902867	6.511530	8.983740
2	11.839800	24.492500	17.411700
3	10.471400	9.290680	10.987700
4	6.381320	6.654020	5.832310
5	4.339480	5.883890	8.326760
6	9.816720	11.838800	3.685560
7	18.020400	16.580700	31.354100
8	5.856240	8.730250	8.505500
9	3.485840	0.946596	2.255210
10	15.646200	14.504600	13.302600
11	3.436060	9.997730	6.463480
12	18.907100	28.830800	12.348500
mean SNR rms	9.09195	12.02184	10.78810
standard deviation	6.02927	8.01670	7.72642
standard error of mean	1.74050	2.31422	2.23042
Grand-average of signals SNR rms	17.22040	23.69880	26.66010
Grand-average of signals SNR dB	24.72086	27.49453	28.51724

	Noisy for F0		
Subjects	Allformants	F1Only	F2&F3Only
1	9.521240	8.496000	3.401530
2	10.403600	9.239560	3.597820
3	2.908920	1.877180	0.430822
4	0.494907	2.276420	1.554660
5	9.958260	7.457040	6.069700
6	6.000630	2.978300	2.649440
7	21.813800	19.585700	5.372940
8	5.071750	4.650410	2.039920
9	4.500970	6.023210	4.760050
10	11.028700	9.960230	2.855030
11	13.066400	11.198200	3.717720
12	5.940140	7.172080	3.132050
	8.39244	7.57619	3.29847
	5.61689	4.85705	1.59229
	1.62146	1.40211	0.45965
	12.91800	11.06020	3.91915
	22.22391	20.87526	11.86384

Table C- 2: rms-SNR of the response at F_0 (all subjects)

Appendix D: Amplitude and SNR of the Responses at F_1

Amplitudes:

	Quiet for F1		
Subjects	Allformants	F1Only	F2&F3Only
1	0.006247	0.002439	0.000493
2	0.006280	0.004449	0.000743
3	0.001988	0.000916	0.001122
4	0.002524	0.002986	0.000110
5	0.008750	0.007195	0.001635
6	0.002249	0.001536	0.000856
7	0.000916	0.011057	0.001968
8	0.004960	0.004588	0.000824
9	0.006366	0.007821	0.002606
10	0.003322	0.002683	0.000955
11	0.004046	0.003007	0.001314
12	0.004970	0.004618	0.001680
rms mean (micro-volts)	0.00438	0.00444	0.00119
standard deviation	0.00229	0.00294	0.00069
standard error of mean	0.00066	0.00085	0.00020
Grand-average of signals (rms mean micro-volts)	0.00129	0.00142	0.00045

	Noisy for F1		
Subjects	Allformants	F1Only	F2&F3Only
1	0.004780	0.002999	0.000657
2	0.002015	0.003523	0.000180
3	0.001243	0.001599	0.001196
4	0.003065	0.002703	0.001233
5	0.005625	0.004144	0.001078
6	0.001107	0.000371	0.000304
7	0.006236	0.004133	0.001244
8	0.001876	0.003040	0.001111
9	0.002624	0.004938	0.000572
10	0.002099	0.000837	0.000563
11	0.001360	0.002117	0.000179
12	0.003118	0.003202	0.000994
	0.00293	0.00280	0.00078
	0.00173	0.00137	0.00041
	0.00050	0.00040	0.00012
	0.00060	0.00085	0.00011

Table D- 1: Amplitude of the response at F_1 (all subjects)

SNRs:

	Quiet for F1		
Subjects	Allformants	F1Only	F2&F3Only
1	5.575010	2.325640	0.518832
2	8.700010	6.640210	1.147670
3	2.029470	0.768534	1.117800
4	2.278370	2.946630	0.110974
5	15.364100	13.877600	2.329460
6	2.203850	3.740450	0.980693
7	1.417150	15.414700	3.261390
8	5.213300	5.084950	1.104690
9	10.309200	8.107760	2.791930
10	6.944210	3.785830	1.665530
11	4.823440	4.445610	1.708820
12	5.644900	6.971570	1.992400
mean SNR rms	5.87525	6.17579	1.56085
standard deviation	4.06302	4.46779	0.92032
standard error of mean	1.17289	1.28974	0.26567
Grand-average of signals SNR rms	4.97987	6.12417	1.91664
Grand-average of signals SNR dB	13.9443	15.74094	5.65081

	Noisy for F1		
Subjects	Allformants	F1Only	F2&F3Only
1	4.560620	2.639190	0.612776
2	2.889290	4.273530	0.243121
3	1.218750	1.661720	1.067380
4	2.637740	2.700460	1.274710
5	8.303820	6.488850	1.410240
6	1.417820	0.729260	0.516674
7	7.261090	4.605770	1.690450
8	2.467190	3.080250	1.051940
9	3.509700	5.443400	0.792997
10	3.173920	1.369890	0.725263
11	0.001360	2.858280	0.248573
12	2.966780	3.665640	0.926942
	3.36734	3.29302	0.88009
	2.38347	1.69508	0.44720
	0.68805	0.48933	0.12910
	2.46260	3.40766	0.48171
	7.82787	10.64913	-6.34432

Table D- 2: rms-SNR of the response at F_1 (all subjects)

Appendix E: Statistical Tests at the Response Frequency (F_0 or F_1) relative to Surrounding Noise

'+-' config:

Legend:

when p<0.001=***

when p<0.01= **

when p<0.05 = *

when p<0.1 and >0.05=+

	p-values					
	Quiet for F0					
Subjects	Allformants	significance	F1Only	significance	F2&F3Only	significance
1	0.447936		0.000000	***	0.000000	***
2	0.000000	***	0.000000	***	0.000000	***
3	0.000000	***	0.000000	***	0.000000	***
4	0.000000	***	0.000000	***	0.000000	***
5	0.000001	***	0.000000	***	0.000000	***
6	0.000000	***	0.000000	***	0.000017	***
7	0.000000	***	0.000000	***	0.000000	***
8	0.000000	***	0.000000	***	0.000000	***
9	0.000044	***	0.414162		0.009469	**
10	0.000000	***	0.000000	***	0.000000	***
11	0.000056	***	0.000000	***	0.000000	***
12	0.000000	***	0.000000	***	0.000000	***
Grand-average of signals	0.000000	***	0.000000	***	0.000000	***

	F-ratio		
	Quiet for F0		
	Allformants	F1Only	F2&F3Only
1	0.820000	42.400000	80.710000
2	140.180000	599.880000	303.170000
3	109.650000	86.320000	120.730000
4	40.720000	44.280000	34.020000
5	18.830000	34.620000	69.330000
6	96.370000	140.160000	13.580000
7	324.740000	274.920000	983.080000
8	34.300000	76.220000	72.340000
9	12.150000	0.900000	5.090000
10	244.800000	210.380000	176.960000
11	11.810000	99.950000	41.780000
12	357.480000	831.220000	152.490000
	296.550000	561.620000	710.770000

Table E- 1: p-values and F-ratios for the response at F_0 (all subjects, quiet)

	Noisy for F0					
	Allformants	significance	F1Only	significance	F2&F3Only	significance
1	0.000000	***	0.000000	***	0.000066	***
2	0.000000	***	0.000000	***	0.000026	***
3	0.000636	***	0.036439	*	0.831127	
4	0.783622		0.008735	**	0.098782	+
5	0.000000	***	0.000000	***	0.000000	***
6	0.000000	***	0.000467	***	0.001951	**
7	0.000000	***	0.000000	***	0.000000	***
8	0.000000	***	0.000000	***	0.020855	*
9	0.000000	***	0.000000	***	0.000000	***
10	0.000000	***	0.000000	***	0.000806	***
11	0.000000	***	0.000000	***	0.000014	***
12	0.000000	***	0.000000	***	0.000232	***
Grand-average of signals	0.000000	***	0.000000	***	0.000005	***

	Noisy for F0		
	Allformants	F1Only	F2&F3Only
1	90.650000	72.180000	11.570000
2	108.230000	85.370000	12.940000
3	8.460000	3.520000	0.190000
4	0.240000	5.180000	2.420000
5	99.170000	55.610000	36.840000
6	36.010000	8.870000	7.020000
7	475.840000	383.600000	28.870000
8	25.720000	21.630000	4.160000
9	20.260000	36.280000	22.660000
10	121.630000	99.210000	8.150000
11	170.730000	125.400000	13.820000
12	35.290000	51.440000	9.810000
	166.870000	122.330000	15.360000

Table E- 2: p-values and F-ratios for the response at F_0 (all subjects, noisy)

-- config:

	p-values					
	Quiet for F1					
	Allformants	significance	F1Only	significance	F2&F3Only	significance
1	0.000000	***	0.005748	**	0.764502	
2	0.000000	***	0.000000	***	0.272088	
3	0.018839	*	0.555717		0.290691	
4	0.007010	**	0.000315	***	0.987762	
5	0.000000	***	0.000000	***	0.005656	**
6	0.009520	**	0.000004	***	0.385411	
7	0.139105		0.000000	***	0.000060	***
8	0.000000	***	0.000000	***	0.299095	
9	0.000000	***	0.000000	***	0.000683	***
10	0.000000	***	0.000003	***	0.066784	+
11	0.000000	***	0.000000	***	0.058124	+
12	0.000000	***	0.000000	***	0.021646	*
Grand-average of signals	0.000000	***	0.000000	***	0.028490	*

	F-ratio		
	Quiet for F1		
	Allformants	F1Only	F2&F3Only
1	31.080000	5.410000	0.270000
2	75.690000	44.090000	1.320000
3	4.120000	0.590000	1.250000
4	5.190000	8.680000	0.010000
5	236.060000	192.590000	5.430000
6	4.860000	13.990000	0.960000
7	2.010000	237.610000	10.640000
8	27.180000	25.860000	1.220000
9	106.280000	65.740000	7.790000
10	48.220000	14.330000	2.770000
11	23.270000	19.760000	2.920000
12	31.860000	48.600000	3.970000
	24.800000	37.500000	3.680000

Table E- 3: p-values and F-ratios for the response at F_1 (all subjects, quiet)

	Noisy for F1					
	Allformants	significance	F1Only	significance	F2&F3Only	significance
1	0.000000	***	0.001418	**	0.687826	
2	0.000421	***	0.000000	***	0.942635	
3	0.230926		0.067593	+	0.323789	
4	0.001428	**	0.001062	**	0.201626	
5	0.000000	***	0.000000	***	0.141752	
6	0.138852		0.589038		0.766204	
7	0.000000	***	0.000000	***	0.061676	+
8	0.003110	**	0.000158	***	0.334344	
9	0.000015	***	0.000000	***	0.535114	
10	0.000096	***	0.157980		0.592439	
11	0.006617	**	0.000492	***	0.940114	
12	0.000285	***	0.000006	***	0.426314	
Grand-average of signals	0.003177	**	0.000027	***	0.53752	

	Noisy for F1		
	Allformants	F1Only	F2&F3Only
1	20.800000	6.970000	0.380000
2	8.350000	18.260000	0.060000
3	1.490000	2.760000	1.140000
4	6.960000	7.290000	1.620000
5	68.950000	42.110000	1.990000
6	2.010000	0.530000	0.270000
7	52.720000	21.210000	2.860000
8	6.090000	9.490000	1.110000
9	12.320000	29.630000	0.630000
10	10.070000	1.880000	0.530000
11	5.250000	8.170000	0.060000
12	8.800000	13.440000	0.860000
	6.06000	11.61000	0.62000

Table E- 4: p-values and F-ratios for the response at F_1 (all subjects, noisy)

Appendix F: Detailed ANOVA and Post-hoc Statistical Test Results

Table F-1 to F-4 presents the F-ratios and *p-values* obtained using the amplitudes and local rms-SNR of the subjects' responses at F_0 and F_1.

ANOVA for response at F0
Sphericity assumed here since Mauchly test >0.05 for both Content and Interaction

Source	F-ratio	P
TypeofStimulus	2.625	0.134
Content	5.503	0.012
TypeofStimulus*Content	8.666	0.002

Table F- 1: ANOVA results for response at F_0 amplitude (+- config)

ANOVA for response at F0
Sphericity assumed here since Mauchly test >0.05 for both Content and Interaction

Source	F	P
TypeofStimulus	6.891	0.024
Content	6.174	0.007
TypeofStimulus*Content	3.846	0.037

Table F- 2: ANOVA results for response at F_0 rms-SNR (+- config)

ANOVA for response at F1
Sphericity assumed here for Content since Mauchly test >0.05 but not for TypeofStimulus*Content (taking Greenhouse-Geisser here)

Source	F	P
TypeofStimulus	18.083	0.001
Content	20.242	0.000
TypeofStimulus*Content	1.244	0.292

Table F- 3: ANOVA results for response at F_1 amplitude (-- config)

ANOVA for response at F1
Sphericity assumed here for Content since Mauchly test >0.05 but not for TypeofStimulus*Content (taking Greenhouse-Geisser here)

Source	F	P
TypeofStimulus	19.002	0.001
Content	16.357	0.000
TypeofStimulus*Content	1.994	0.181

Table F- 4: ANOVA results for response at F_1 rms-SNR (--config)

Table F-5 to F-8 presents the p-values for the two-tailed repeated measures t-tests using SPSS for the response at F_0 and F_1:

T-tests of mean rms amplitude

Q1: When the signal is clean, do Allformants, F1Only, F2&F3Only contribute differently to the response?

Q1	
Allformantsclean –F1_onlyclean	0.207
Allformantsclean - F2_F3_onlyclean	0.325
F1_onlyclean - F2_F3_onlyclean	0.918

Q2: When the signal is noisy, do Allformants, F1Only, F2&F3Only contribute differently to the response?

Q2:	
Allformants_noisy - F1_onlynoisy	0.167
Allformants_noisy - F2_F3_onlynoisy	0.000
F1_onlynoisy -F2_F3_onlynoisy	0.000

Q3a) Does noise changes the response with the Allformants stimulus?
Q3b) Does noise changes the response with the F1Only stimulus?
Q3c) Does noise changes the response with the F2&F3Only stimulus?

Q3 a)	
Allformantsclean - Allformants_noisy	0.842
Q3b)	
F1_onlyclean - F1_onlynoisy	0.259
Q3c)	
F2_F3_onlyclean - F2_F3_onlynoisy	0.003

Table F- 5: Two-tailed t-tests results for response at F_0 (amplitude)

T-tests for mean rms-SNR

Q1	
Allformantsclean -F1_onlyclean	0.059
Allformantsclean - F2_F3_onlyclean	0.326
F1_onlyclean - F2_F3_onlyclean	0.580

Q2	
Allformants_noisy - F1_onlynoisy	0.102
Allformants_noisy - F2_F3_onlynoisy	0.003
F1_onlynoisy -F2_F3_onlynoisy	0.003

Q3 a)	
Allformantsclean - Allformants_noisy	0.728
Q3b)	
F1_onlyclean - F1_onlynoisy	0.079
Q3c)	
F2_F3_onlyclean - F2_F3_onlynoisy	0.005

Table F- 6: Two-tailed t-tests results for response at F_0 (rms-SNR)

T-tests for mean rms amplitude

Q1	
Allformantsclean -F1_onlyclean	0.955
Allformantsclean - F2_F3_onlyclean	0.000
F1_onlyclean - F2_F3_onlyclean	0.001

Q2	
Allformants_noisy – F1_onlynoisy	0.756
Allformants_noisy - F2_F3_onlynoisy	0.001
F1_onlynoisy -F2_F3_onlynoisy	0.000

Q3 a)	
Allformantsclean - Allformants_noisy	0.072
Q3b)	
F1_onlyclean - F1_onlynoisy	0.017
Q3c)	
F2_F3_onlyclean - F2_F3_onlynoisy	0.094

Table F- 7: Two-tailed t-tests results for response at F_1 (amplitude)

Q1	
Allformantsclean -F1_onlyclean	0.825
Allformantsclean -F2_F3_onlyclean	0.003
F1_onlyclean - F2_F3_onlyclean	0.001

Q2	
Allformants_noisy - F1_onlynoisy	0.884
Allformants_noisy - F2_F3_onlynoisy	0.002
F1_onlynoisy -F2_F3_onlynoisy	0.000

Q3 a)	
Allformantsclean - Allformants_noisy	0.035
Q3b)	
F1_onlyclean - F1_onlynoisy	0.011
Q3c)	
F2_F3_onlyclean - F2_F3_onlynoisy	0.020

Table F- 8: Two-tailed t-tests results for response at F_1 (rms-SNR)

Appendix G: Informed Consent Form

Measurement of Speech-Evoked Potentials in Normal Hearing Subjects

You are invited to participate in this study. This informed consent form is to make sure that you understand the purpose of this study and the nature of your involvement. The research project is being conducted by Dr. Hilmi Dajani, Assistant Professor at the School of Information Technology and Engineering of the University of Ottawa (tel: 613-562-5800, ext. 6217).

Purpose of the study: The purpose of this study is to investigate the brain's electrical response to speech stimuli. This electrical signal, called the auditory-evoked potential, allows clinicians to diagnose hearing problems and localize their origin in the brain. The sounds that clinicians most commonly use for this test are brief clicks or other artificial sounds. These types of sounds do not provide sufficient information to assess the ability to hear speech. As a result, we are interested in developing tests that record the brain's electrical response to speech, or the Speech-Evoked Potential, with a diverse set of speech samples and in different acoustic environments. The knowledge gained is expected to help in the development of better diagnostic tests of hearing loss and tools for evaluating hearing aids.

Procedure: First of all, we will start by measuring the activity from your brainstem, using small electrodes gently placed on your scalp and earlobes. You will then listen to sounds at a comfortable level for the duration of approximately 90 minutes.

Subject description: Adults with normal hearing and without any external or middle ear disease will be included.

Risks to participating: There is no danger or risk to health associated with this study. All procedures have been pre-tested and they have been used routinely for many years in hospitals, clinics, and laboratories.

Informed Consent Form (continued)

Withdrawing from the study: Your participation in this study is voluntary. You may withdraw from the study at any time, by verbally informing any of the investigators, even after signing this form.

If you have any concerns with regards to the ethical conduct of the study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, 550 Cumberland Street, Room 159, Ottawa, ON K1N 6N5, tel.: 613-562-5841, email: ethics@uottawa.ca.

Compensation: You will not receive monetary compensation for this study.

Confidentiality: Any information about you collected during this study will be kept strictly confidential. Your name will not be associated with the collected data in any way. While the results will appear in students' dissertations and may also be published, you will not be identified.

The data collection will be conducted by Dr. Dajani, his graduate student, or his research assistant. The records will be kept on computer files. The files will be password-protected. The data will be kept in Dr. Dajani's office when not in use and will be conserved for a maximum period of 10 years, after which the computer files will be deleted.

In closing: With your participation, you will be given a copy of this consent form. At the conclusion of the study, should you wish, you will be provided with a summary of the results. You may ask questions at any time, even after signing this consent form.

Signatures: I have read the above description of the study and understand the conditions of participation. My signature indicates that I agree to participate in the study.

Name of the participant (please print):

_____

Participant's signature:

_____ **Date:** _____

Researcher's signature:

_____ **Date:** _____

References

Aiken, S.J., Picton, T.W. 2006. Envelope following responses to natural vowels Audiology and Neurotology 11, 213-232.

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

Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H. 2008a. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clinical Neurophysiology 119, 922-933.

Akhoun, I., Moulin, A., Jeanvoine, A., Ménard, M., Buret, F., Vollaire, C., Scorretti, R., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H. 2008b. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: An experimental parametric study. Journal of Neuroscience Methods 172, 196-205.

Anderson, S., Skoe, E., Chandrasekaran, B., Kraus, N. 2010. Neural Timing Is Linked to Speech Perception in Noise. The Journal of Neuroscience 30, 4922-4926.

Assmann, P., Summerfield, Q. 2004. Chapter 5: Perception of Speech Under Averse Conditions. In: Popper, A.N., Fay, R.R., (Eds.), Speech Processing in the Auditory System, Vol. 18, 1 ed. Springer New York, New York. pp. 231-309.

Bartlett, E.L., Wang, X. 2007. Neural representations of temporally modulated signals in the auditory thalamus of awake primates. Journal of Neurophysiology 97, 1005-1017.

Bear, F.M., Connors, W.B., Paradiso, A.M. 2007. Chapter 11: The Auditory and Vestibular Systems, Neuroscience, 3 ed. Lippincott Williams & Wilkins, Philadelphia. pp. 343-387.

Bledscoe, S.C., Jr., Moushegian, G. 1980. The 500 Hz frequency-following potential in kangaroo rat: an evaluation with noise masking. Electroencephalography and Clinical Neurophysiology 48, 654-663.

Burkard, R.F., Eggermont, J.J., Don, M. 2007. Section 1: What are auditory evoked potentials?, Auditory evoked potentials: basic principles and clinical application, 1 ed. Lippincott Williams & Wilkins, Philadelphia pp. 7-20.

Cariani, P.A., Delgutte, B. 1996. Neural Correlates of the Pitch of Complex Tones. I. Pitch and Pitch Salience. Journal of Neurophysiology 76, 1698-1716.

Carlyon, R.P. 1996. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker Journal of the Acoustical Society of America 99, 517-524.

Carlyon, R.P., Shackleton, T.M. 1994. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? Journal of the Acoustical Society of America 95, 3541-3554.

Cedolin, L., Delgutte, B. 2004. Representations of the pitch of complex tones in the auditory nerve. In: Pressnitzer D., de Cheveigne A., McAdams S., Collet L., (Eds.), Auditory Signal Processing: Physiology, Psychoacoustics, and Models. Springer Verlag. pp. 107-116.

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

Culler, E.A., Coakley, R.D., Lowy, K., Gross, N. 1943. A revised frequency-map of the guinea-pig cochlea. The American Journal of Psychology 56, 475-500.
Cunningham, J., Nicol, G.T., Zecker, S.G., Bradlow, A.R., Kraus, N. 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. Clinical Neurophysiology 112, 758-767.
Cunningham, J., Nicol, G.T., King, C., Zecker, S.G., Kraus, N. 2002. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. Hearing Research 169, 97-101.
Dajani, R.H., Purcell, D., Wong, W., Kunov, H., Picton, T.W. 2005. Recording Human Evoked Potentials That Follow the Pitch Contour of a Natural Vowel. IEEE Transactions on Biomedical Engineering 52, 1614-1618.
Darwin, C.J., McKeown, J.D., Kirby, D. 1989. Perceptual compensation for transmission channel and speaker effects on vowel quality. Speech Communication 8, 221-234.
Deutsch, S., Micheli-Tzanakou, E. 1987. 12.4 Digital Signal Processing of Evoked Potentials, Neuroelectric Systems. New York University Press New York, USA. pp. 450-462.
Don, M., Elberling, C. 1994. Evaluating residual background noise in human auditory brain-stem responses. Journal of the Acoustical Society of America 96, 2746-2757.
Dutoit, T. 1997. Chapter 7: Synthesis Strategies, An introduction to text-to-speech synthesis, 1 ed. Kluwer Academic Publishers, Netherlands. pp. 177-201.
Eduardo, R.M. 2002. Chapter 3: Spectrum Modelling Approaches: from additive to analysis-resynthesis and formant, Computer Sound Design: Synthesis techniques and programming, 2 ed. Focal Press, Woburn (USA). pp. 49-67.
Field, A. 2009. Repeated-measures designs (GLM 4), Discovering statistics using SPSS, 3rd ed. SAGE Publications Ltd., London, UK. pp. 457-505.
Frisina, R.D., Walton, J.P., Karcich, K.J. 1994. Dorsal cochlear nucleus single neurons can enhance temporal processing capabilities in background noise Experimental Brain Research 102, 160-164.
Frisina, R.D., Karcich, K.J., Tracy, T.C., Sullivan, D.M., Walton, J.P., Columbo, J. 1996. Preservation of amplitude modulation coding in the presence of background noise by chinchilla auditory-nerve fibers. Journal of the Acoustical Society of America 99, 475-490
Galbraith, G.C. 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalography and clinical Neurophysiology 92, 321-330.
Geisler, C.D., Greenberg, S. 1986. A two-stage nonlinear cochlear model possesses automatic gain control. Journal of the Acoustical Society of America 80, 1359-1363.
Gelfand, S. 2001. Chapter 2: Anatomy and Physiology of the Auditory System, Essentials of Audiology, 2 ed. Thieme Medical Publishers, New York. pp. 37-91.
Gockel, H., Carlyon, R.P. 2004. Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms. Journal of the Acoustical Society of America 116, 1092-1104.
Goldstein, B.E. 2009. Chapter 11: Sound, the Auditory System, and pitch perception, Sensation and perception, 8 ed. Cengage Learning, Belmont (USA). pp. 259-291.
Greenberg, S., Ainsworth, W.A. 2004. Chapter 1: Speech Processing in the Auditory System: An Overview. In: Popper, A.N., Fay, R.R., (Eds.), Speech Processing in the Auditory System, Vol. 18, 1 ed. Springer New York, New York. pp. 1-63.

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

Greenberg, S., Popper, A.N., Ainsworth, W.A., Fay, R.R. 2004a. Chapter 4: Physiological Representation of Speech, Speech Processing in the Auditory System, Vol. 18, 1 ed. Spinger New York, New York.

Greenberg, S., Ainsworth, W.A., Popper, A.N., Fay, R.R. 2004b. Chapter 5: Perception of Speech Under Averse Conditions, Speech Processing in the Auditory System, Vol. 18, 1 ed. Springer New York, New York. pp. 231-309.

Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., Patterson, R.D. 2001. Encoding of the temporal regularity of sound in the human brainstem. Nature Neuroscience 4, 633-637.

Grimault, N., Micheyl, C., Carlyon, R.P., Collet, L. 2002. Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. Perception & Psychophysics 64, 189-197.

Hall, D.A., Plack, C.J. 2008. Pitch Processing Sites in the Human Auditory Brain. Cerebral Cortex 19, 576-585.

Hall, W.J., III. 2007a. Chapter 8: Frequency-specific Auditory Brainstem Response (ABR) and Auditory Steady-State Response (ASSR)," New Handbook of Auditory Evoked Responses, 2 ed. Allyn & Bacon, Columbus(USA). pp. 258-313.

Hall, W.J., III. 2007b. Chapter 3: Introduction to auditory evoked response measurement," in New Handbook of Auditory Evoked Responses, New Handbook of Auditory Evoked Responses, 1 ed. Allyn & Bacon:Boston, Boston. pp. 58-109.

Hall, W.J., III. 2007c. Chapter 7: ABR Analysis and Interpretation, New Handbook of Auditory Evoked Responses. Allyn & Bacon, Columbus(USA). pp. 212-257.

Hall, W.J., III. 2007d. Chapter 1: Overview of Auditory Neurophysiology: Past, Present, and Future,, New Handbook of Auditory Evoked Responses, 1 ed. Allyn & Bacon, Boston. pp. 1-35.

Harpaz, Y., Levkovitz, Y., Lavidor, M. 2009. Lexical ambiguity resolution in Wernicke's area and its right homologue. Cortex 45, 1097-1103.

Holmes, J., Holmes, W. 2001. Chapter 5: Message Synthesis from Stored Human Speech Components, Speech synthesis and recognition, 2 ed. Taylor & Francis, London (UK). pp. 67-81.

Houtsma, A.J.M., Smurzynski, J. 1990. Pitch identification and discrimination for complex tones with many harmonics Journal of the Acoustical Society of America 87, 304-310.

Johnson, K.L., Nicol, G.T., Kraus, N. 2005. Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear & Hearing 26, 424-434.

Johnson, K.L., Nicol, G.T., Zecker, S.G., Bradlow, A.R., Skoe, E., Kraus, N. 2008. Brainstem encoding of voiced consonant–vowel stop syllables. Clinical Neurophysiology 119, 2623-2635.

Klatt, H.D. 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971-995.

Kraus, N., Cheour, M. 2000. Speech Sound Representation in the Brain. Audiology & Neurotology 5, 140-150.

Kraus, N., Nicol, G.T. 2005. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. TRENDS in Neurosciences 28, 176-181.

Kraus, N., McGee, T., Carrell, T.D., Sharma, A. 1995a. Neurophysiologic Bases of Speech Discrimination. Ear & Hearing 16, 19-37.

Kraus, N., McGee, T., Carrell, T., Sharma, A. 1995b. Neurophysiologic Bases of Speech Discrimination. Ear & Hearing 16, 19-37.

Kraus, N., McGee, T., Ferre, J., Hoeppner, J., Carrell, T.D., Sharma, A., Nicol, G.T. 1993. Mismatch Negativity in the Neurophysiologic/Behavioral Evaluation of Auditory Processing Deficits: A Case Study. Ear & Hearing 14, 223-234.

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

Krishnan, A., Gandour, J.T. 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns Brain & Language 110, 135-148.

Krishnan, A., Gandour, J.T., Bidelman, G.M. 2010. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Research 1313, 124-133.

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

Leonardo, B.F., Moraes, J.E., Miranda de Sa, A.M., Yeha, H.C., Moraes, M.F. 2005. Avoiding spectral leakage in objective detection of auditory steady-state evoked responses in the inferior colliculus of rat using coherence. Journal of Neuroscience Methods 144, 249-255.

Lins, G.O., Picton, W.T. 1995. Auditory steady-state responses to multiple simultaneous stimuli. Electroencephalography and clinical Neurophysiology 96, 420-432.

Martin, B.A., Tremblay, K.L., Korczak, P. 2008. Speech Evoked Potentials: From the Laboratory to the Clinic. Ear & Hearing 29, 285-313.

Maurer, K., Schafer, E., Leitner, H. 1980. The effect of varying stimulus polarity (rarefaction vs. condensation) on early auditory evoked potentials (EAEPs). Electroencephalography and Clinical Neurophysiology 50, 332-334.

Micheyl, C., Oxenham, A.J. 2004. Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms. Journal of the Acoustical Society of America 116, 3038-3050.

Møller, A.R. 2003. Sensory Systems: Anatomy and Physiology. 1 ed. Elsevier Science, London (UK).

Møller, A.R. 2006a. Section II: The Auditory Nervous System, Hearing: Anatomy, physiology, and disorders of the auditory system, 2 ed. Elsevier Science London(UK). pp. 75-192.

Møller, A.R. 2006b. Section 1: The Ear, Hearing: Anatomy, physiology, and disorders of the auditory system. Elsevier Science, London UK. pp. p.3-68.

Norman, R.G., Streiner, L.D. 2008. Chapter eight: More than two groups, Biostatistics: The Bare Essentials, 3rd ed. People's Medical Publishing House, Shelton (USA). pp. 72-90.

Otavio, G.L., Picton, T.W. 1995. Auditory steady-state responses to multiple simultaneous stimuli. Electroencephalography and clinical Neurophysiology 96, 420-432.

Palmer, A., Shamma, A.S. 2004. Chapter 4: Physiological Representation of Speech. In: Popper, A.N., Fay, R.R., (Eds.), Speech Processing in the Auditory System, Vol. 18, 1 ed. Spinger New York, New York.

Parbery-Clark, A., Skoe, E., Kraus, N. 2009. Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound. The Journal of Neuroscience 29, 14100-14107.

Penagos, H., Melcher, J.R., Oxenham, A.J. 2004. A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. The Journal of Neuroscience 24, 6810-6815.

Peterson, E.G., Barney, L.H. 1952. Control Methods Used in a Study of the Vowels. The Journal of the Acoustical Society of America 24, 175-184.

Picton, T.W., John, M.S. 2004. Avoiding electromagnetic artifacts when recording auditory steady-state responses. Journal of the American Academy of Audiology 15, 541-554.

Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D. 2003. Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. International Journal of Audiology 42, 177-219.

Pittman, L.A., Wiley, L.T. 2001. Recognition of Speech Produced in Noise. Journal of Speech, Language and Hearing Research 44, 487-496.

Rhode, S.W., Greenberg, S. 1994. Encoding of Amplitude Modulation in the Cochlear Nucleus of the Cat. Journal of Neurophysiology 71, 1797-1825.

Rose, E.J., Brugge, F.J., Anderson, J.D., Hind, E.J. 1967. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. Journal of Neurophysiology 30, 769-793.

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

Sachs, M.B., Young, E.D. 1979. Representation of steady-state vowels in the temporal aspects of the discharge patterns of auditory-nerve fibers. Journal of the Acoustical Society of America 66, 1381-1403.

Sanchez, J.T., Gans, D. 2006. Effects of Artifact Rejection and Bayesian Weighting on the Auditory Brainstem Response During Quiet and Active Behavioral Conditions. American Journal of Audiology 15, 154-163.

Secker-Walker, H.E., Campbell, S.L. 1990. Time-domain analysis of auditory-nerve-fiber firing rates. Journal of the Acoustical Society of America 88, 1427-1436.

Shackleton, T.M., Carlyon, R.P. 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. Journal of the Acoustical Society of America 95, 3529-3540.

Skoe, E., Kraus, N. 2010. Auditory Brain Stem Response to Complex Sounds: A Tutorial. Ear & Hearing 31, 302-324.

Song, J.H., Banai, K., Russo, N.M., Kraus, N. 2006. On the relationship between speech- and nonspeech-evoked auditory brainstem responses Audiology and Neurotology 11, 233-241.

Van Santen, P.H.J., Sproat, R., Olive, J., Hirschberg, J. 1997. Section IV: Concatenative Synthesis and Automated Segmentation, Progress in Speech Synthesis, 1 ed. Springer-Verlag, New York. pp. 259-325.

Webster, G.J., Clark, J.W. 1997. Chapter 14: Electrical safety, Medical Instrumentation: Application and Design, 3 ed. Wiley & Sons inc, Danvers (USA). pp. 623-659.

White, L.J., Plack, C.J. 1998. Temporal processing of the pitch of complex tones. Journal of the Acoustical Society of America 103, 2051-2063.

Wible, B., Nicol, G.T., Kraus, N. 2004. A typical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. Biological Psychology 67, 299-317.

Winkler, I., Tervaniemi, M., Näätänen, R. 1997. Two separate codes for missing-fundamental pitch in the human auditory cortex. Journal of the Acoustical Society of America 102, 1072-1082.

Xu, Y., Gandour, J.T., Krishnan, A. 2006. Specificity of experience-dependent pitch representation in the brainstem Neuroreport 17, 1601-1605.

Zhang X., Hood C. D., Chen S. C., Hong E. J. 2002. A signal-to-noise analysis of multifocal VEP responses: an objective definition for poor records. Documenta Ophthalmologica 104, 287-302.