2009-05-28

Frequency Response of the Skin on the Head and Neck During Production of Selected Speech Sounds

Jacob B. Munger
Brigham Young University - Provo

FREQUENCY RESPONSE OF THE SKIN ON THE HEAD AND NECK DURING PRODUCTION OF SELECTED SPEECH SOUNDS

by

Jacob B. Munger

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Mechanical Engineering

Brigham Young University

August 2009

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Jacob B. Munger

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Scott L. Thomson, Chair
Date	Christopher D. Dromey
Date	Jonathan D. Blotter

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Jacob B. Munger in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Scott L. Thomson
Chair, Graduate Committee

Accepted for the Department

Larry L. Howell
Graduate Coordinator

Accepted for the College

Alan R. Parkinson
Dean, Ira A. Fulton College of Engineering and Technology

ABSTRACT

FREQUENCY RESPONSE OF THE SKIN ON THE HEAD AND NECK
DURING PRODUCTION OF SELECTED SPEECH SOUNDS

Jacob B. Munger

Department of Mechanical Engineering

Master of Science

Vibration within the vocal tract during speech is transmitted through body tissue to the skin surface and can be used to transmit speech. Achieving quality speech signals using skin vibration is desirable but problematic, primarily due to the different sources of sound during speech. The objective of this study was to characterize the frequency content of speech signals at various locations on the head and neck. Signals were recorded using accelerometers attached to 15 locations on the heads and necks of 14 males and 10 females as well as a microphone to record audible speech. The subjects produced several isolated phonemes and one phrase. The power spectral densities (PSDs) of the phonemes were used to determine a quality ranking for each location and sound. A spectrogram of the phrase was used to compare the response at selected locations. A perceptual listening test was conducted and compared to the PSD rankings. The PSD

rankings were also calculated for signals recorded with background noise in order to identify locations that are least sensitive to external noise. With background noise, the frequency response of the skin was also used to study how the skin itself responds to external noise. The signal-to-noise ratio (SNR) was found for various sounds and locations with and without the presence of background noise. The frequency response of a concentrated area of the neck was also studied.

Notably, while high frequency content was found to be attenuated at locations on the throat near the thyroid cartilage, it was detectable at some other locations. The best locations for speech transmission were found to be generally common to males and females in quiet environments but varied with background noise. During speech in the presence of background noise, the accelerometers performed better than the microphone when compared to the PSD of a clean microphone recording of the same sound. All SNR of all locations were influenced somewhat by external noise, some considerably more than others. Some neck locations may be better suited for contact microphone placement other than directly over the thyroid cartilage (where many commercial contact microphones are currently worn) if the neck is the preferred location for contact microphone placement.

ACKNOWLEDGMENTS

I would like thank my advisor, Dr. Scott Thomson, for all the help and assistance he has given me over the last four years that I have been working with him. He has been very generous in financially supporting my graduate research and studies. I would also like to thank him for all the advice he has given me as I have planned my academic studies as well as in the research we have conducted together. I have enjoyed learning from his experience and knowledge.

I would also like to thank the rest of my thesis committee, Dr. Christopher Dromey and Dr. Jonathan Blotter, for their assistance in improving this thesis. I would also like to thank Dr. Dromey for the use of his sound booth and perceptual rating software, as well as the advice he has given me throughout my research.

I am thankful for all of the volunteer subjects, and those who help in the data collection process for giving of their time to assist me in my research. I also acknowledge support from the Rocky Mountain NASA Space Grant Consortium Fellowship. Early support by Tri-Mast International is also acknowledged.

I am grateful for my parents and the loving support and encouragement they have given me throughout my education. I am also thankful for my beautiful bride to be, Bridget, for the shinning light and motivation she is to me. I love her and the wonderful example she is in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This chapter introduces the thesis topic, presents necessary background information, and outlines the remainder of the thesis.

1.1 Speech Formation and Transmission

Voiced speech sounds are produced when pressurized air from the lungs passes through the larynx, causing the vocal folds to vibrate. Acoustic vibration resulting from glottal airflow pressure fluctuations and from structural (tissue) vibration is transmitted through the vocal tract (including the oral and nasal cavities). Vocal fold vibration provides the source for all voiced sounds; however, each sound is formed differently by the position of the lips and tongue.. Vowel sounds (e.g., /æ/; see Table A-1 in Appendix A which describes the sounds used in this thesis) are shaped in the mouth using the lips and tongue and are transmitted primarily through the open mouth. Nasal sounds (e.g., /n/ and /m/) are also influenced by mouth shape, but no air flows through the mouth; thus the sound is primarily transmitted through the nasal passages.

For some speech sounds such as unvoiced fricatives and stops, the vocal folds do not vibrate. Unvoiced sounds have a variety of sound sources, such as turbulence, aspects of which can be more complex than voiced sounds.

1.2 Contact Microphones

The vibration resulting from speech is not only transmitted through the vocal tract but also through body tissue surrounding the vocal tract to the skin surface. The skin surface vibration can be sensed by transducers placed on the skin, often called contact microphones, and can be used to transmit speech in communication systems. As opposed to acoustic microphones which sense speech signals via airborn pressure fluctuations, contact microphones on the skin have the potential to sense little background noise from the surrounding environment. As a result, contact microphones have the potential to enhance speech clarity in elevated ambient noise environments. This presents significant potential benefits to military, aerospace, and general purpose cell phone and two-way radio communications. However, a drawback to most contact microphones designs is that the intelligibility of the signal sensed on the skin is significantly worse than that of a signal sensed by an acoustic microphone.

1.3 Thesis Research Overview

Lacking in previous studies are detailed reports of the frequency response of head and neck skin vibration during speech production at various anatomical locations, as well as the frequency analysis of sounds other than vowels at numerous locations on the head and neck. The purpose of the research presented in this thesis is to assist in the development of improved contact microphones and associated signal processing methods by contributing to an improved understanding of the frequency response of the skin at multiple locations on the head and neck during speech production of various phonemes.

Skin vibration signals were collected at various locations on the head and neck in order to gain a better understanding of the speech signal that is transmitted to each location. To collect the skin vibration signal, accelerometers were attached to 15 locations on the head and neck of 24 human subjects. A microphone was used to record the audible speech signal. The subjects spoke a phrase and various voiced and unvoiced phonemes with and without the presence of background noise. The accelerometer and microphone power spectral densities (PSDs) were compared and the accelerometer locations were ranked according to how well their signal's PSD corresponded to that of the microphone. A juried listening test was conducted to compare the PSD rankings to perceptual rankings. Here a spectrogram of a spoken phrase was also used to compare the amplitudes of various frequencies at selected locations.

The PSD rankings were also calculated for signals recorded with background noise in order to identify locations that are least sensitive to external noise. With background noise, the frequency response of the skin when the person was not speaking was used to determine how the skin itself responds to external noise. The signal-to-noise ratio (SNR) was found for various sounds and at various locations with and without the presence of background noise to determine sensitivity to external noise. The frequency response of 12 locations on the neck was also studied to identify preferred locations in that region.

1.4 Thesis Outline

Chapter 2 is a review of previous research relating to contact microphone performance and skin vibration signals. Chapter 3 describes the experimental setup and data collection methods used in this research. Also presented are the results and discussion of the PSD analysis and rankings, spectrogram analysis, perceptual ratings, and signal-to-noise analysis. Chapter 4 presents the results of a PSD analysis with the addition of background noise. Chapter 5 presents the SNR analysis and results. Chapter 6 describes the experimental setup, data collection methods, and results for the PSD analysis of a concentrated area on the neck. Chapter 7 contains the conclusions reached as a result of this work and recommendations for future work.

2 Literature Review

This chapter reviews prior work relating to contact microphones and skin vibration signals. First, early work done in the 1950s and 1960s is reviewed. These studies show the potential of contact microphones to transmit speech but also identify some of the inherent limitations of contact microphones. Although these early studies provide useful information, they are limited by the hardware and signal processing capabilities of that era. Surprisingly, even with advances in measurement technology, it is shown that there have not been many investigations into the potential of using skin surface vibration to transmit speech using modern hardware and signal processing techniques.

Secondly, a study by the United States Army Aeromedical Research Laboratory (Acker-Mills et al., 2004) that serves as a starting point and, in many ways, a motivation for this thesis research is reviewed.

Thirdly, studies that use accelerometers to study skin surface vibration for purposes other than speech transmission are reviewed. In these studies the skin surface vibration has been used to help understand the dynamic characteristics of the human skull, record tracheal sounds in patients with sleep apnea, estimate the radiated sound pressure level of audible speech, and to monitor voice use.

Fourth, studies that combine contact microphones and acoustic microphones to improve speech quality in high noise environments are reviewed. Finally, studies that measure the transfer function of the neck and studies that use signal processing techniques to improve speech quality of contact microphones are reviewed. This chapter ends with a brief summary of the prior works reviewed, an outline of some their limitations, and a discussion of how this thesis research addresses some of these limitations.

2.1 Early Work

The potential for contact microphones to provide clear speech communication in high noise environments prompted early researchers to investigate the speech signal that was detected on the skin. Moser and Oyer (1958) conducted a study to measure the intensity of 12 vowel sounds at 16 locations on the face and neck of 3 male subjects. Measurements were made using a bone oscillator (a device that detects bone-conducted speech), one location at a time. Addressing the concern that different transducer pressures on the skin would affect the results, they found that varying the transducer pressure only slightly affected the signal. This study only presented the average intensity levels for each of the sounds and locations; no spectrum comparisons where made. When the recordings were subjectively compared by the researchers, they judged the forehead to yield the most faithful representation of the sounds while the locations on the neck yielded the highest intensity signals. The main conclusion of the study was that the relative intensity of the sound measured on the skin surface decreases with distance from the larynx.

Snidecor et al. (1959) conducted a study to identify locations on the face and neck that yield good intelligibility and quality. A contact microphone was placed at eight locations on the head and neck of one male speaker while he produced four vowel sounds and a speech sample. This study found the intensity levels to be comparable to those of the Moser and Oyer (1958) study. No spectral analysis was performed. The authors themselves judged the intelligibility of the speech samples based on a scale of poor, fair, average, good, and excellent. They found that the mandible, chin, nose, mastoid process, forehead, and ear canal had good intelligibility, while the larynx and zygoma were average. They surveyed 24 college students for their preference of the speech sample at each location using the method of (A-B) paired comparison and forced preference choice. The preference study results are shown in Table 2-1.

Table 2-2-1 Paired comparison preference results for the Sidecor et al. (1959) study.

Position	Total Preferences
Forehead	140
Mastoid	93
Larynx	93
Seventh cervical vertebra	90
Zygomatic Arch	70
Tempro-mandibular joint	69
Mandible	63
Nose	50

Although results for intelligibility in noise were not presented, the authors noted that the contact microphone they used had a low SNR but still yielded an intelligible signal at some locations in the presence of noise. The authors recommended the forehead, mastoid process, larynx, mandibular angle, ear canal, and nose as possible suitable locations for contact microphones.

An early application of contact microphones is found in a report by Hayes and Meltzer (1967). In this report the authors explore the use of a contact microphone placed at the forehead to record speech in conversational settings. The contact microphone was attached to the forehead of a subject using a one inch headband which applied 0.2 to 0.3 pounds per square inch of pressure. To test the sensitivity of the contact microphone to external speech noise, the sound intensity level from a contact microphone worn by a listener and of an acoustic microphone, both located six feet from a speaker, were compared. They found that the contact microphone worn by the listener was insensitive to the speaker's speech while the acoustic microphone was very sensitive.

Hayes and Meltzer (1967) listed a few limitations to the use of contact microphones in their report. The recordings made with contact microphones had low fidelity when compared to conventional microphones. It was noted that contact microphones sensed non-speech-related noise from the speaker's mouth, such as chewing gum, gnawing on a pencil, or even scratching their head. The authors also noted that effects of anatomical differences such as skull thickness or clogged nasal passages on the recordings were not known. A limitation to this report is that the contact microphone only had a usable frequency range of 225-2500 Hz, which would have affected the fidelity of the recordings.

2.2 United States Army Study

In 2004 the United States Army Aeromedical Research Laboratory conducted a study (Acker-Mills et al., 2004) that compared noise canceling boom-mounted acoustic microphones to a commercially available throat microphone similar to ones used by Navy SEALS and Army Special Forces. The purpose of the study was to determine if using throat microphones in helicopters would increase speech intelligibility. Data were collected in a reverberation chamber with 90 dB and 106 dB broadband noise to simulate helicopter noise. Talkers wore the throat microphone at a comfortable position and with a pressure of 200 grams of force. An intelligibility study was performed using the Modified Rhyme Test following ANSI specifications.

They found that although the throat microphones had a 10 dB higher signal-to-noise ratio, the signal detected at the throat degraded high frequency content, which in turn degraded intelligibility. It was concluded that the loss of consonant information in the speech signal detected at the throat contributed to reduced speech intelligibility. They also found the throat microphone performed worse than a noise canceling acoustic microphone in high levels of noise.

The conclusion of the study states:

> The current study demonstrates that the use of a throat microphone in noisy environments similar to that of rotary-wing aircraft does not increase speech intelligibility. Thus, it is recommended at this time, the U.S. Army not consider the use of throat microphones in noisy environments. It is possible that future technology will improve throat microphone performance, but <u>consonant information will never be able to be transmitted adequately if speech information is picked up only from the throat area</u>. (underline added)

The last sentence prompts the question: ‘If consonant information cannot be adequately sensed at the throat, can it be sensed at other locations on the face or neck?’ This study served as a motivation and starting point for the research presented in this thesis.

2.3 Accelerometer Use in Studying Skin Vibration

Due to their light weight, favorable frequency range, and sensitivity characteristics, accelerometers have been used to detect small amplitude vibration on the skin surface for purposes other than communication. Cheyne (1993) and Cheyne et al. (2003) used accelerometers to quantify vocal function and monitor voice use. In Cheyne et al. (2003) the accelerometers were attached to the skin using Skin-Bond® adhesive. They found that the first two harmonic peaks for the softest spoken /æ/ were 50 dB above the noise floor. They reported that the accelerometers used were suitable for detecting even the softest phonations tested. They reported that the accelerometers’ dynamic range was sufficient for detecting and recording phonation corresponding to a sound pressure level (SPL) range from 46 to 105 dB (the distance was unreported).

Horáček et al. (2004) used an accelerometer placed on the forehead of a speaker to help understand the dynamic characteristics of the human skull and to verify a computational finite element model of the sound transmission chain in the human skull. The accelerometer was pressed against the forehead using a rubber strap worn around the speaker’s head. At the forehead they found that accelerometer signal frequencies greater than 3500 Hz were significantly more attenuated than those of the acoustic signal. They also found that although high frequency content was attenuated, the forehead signal still had good intelligibility.

Svec et al. (2005) used accelerometers placed near the sternal notch to estimate speech intensity. The accelerometers were attached using surgical adhesive with a suture strip over the surface of the accelerometer. In Pasterkamp et al. (1996) accelerometers were used to record tracheal sounds in patients with sleep apnea. In this study the accelerometers were attached to the skin using double-sided adhesive tape.

2.4 Combination of Acoustic and Contact Microphones

To improve speech transmission in high noise environments, researchers have investigated using combinations of acoustic and contact microphones. Many studies have shown that such combinations can improve word recognition in speech recognition software (Graciarena et al., 2003; Zheng et al., 2003; Dupont et al., 2004; Zhang et al., 2004). However, in these studies the contact microphones were not used to transmit the speech signal itself, but rather as a "trigger" to prevent background noise transmission when the speaker was not talking.

2.5 Transfer Functions

In a few studies, investigators have used transfer function estimates to better understand the frequency response characteristics of the head and neck. The frequency response of the neck was studied using an external shaker to induce vibration on the skin in order to identify a transfer function for use with electrolarynx devices (Norton et al., 1993; Meltzner et al., 2003). Wodicka and Shannon (1990) found a transfer function for the subglottal respiratory system when sound was introduced at the mouth for use in detecting lung disease. Shimamura and Tamiya (2005) presented a method for developing a finite impulse response reconstruction filter for a bone conduction microphone placed at the top of the head. They found that high frequency content was restored using the filter, which increased the clarity of bone-conducted speech.

2.6 Literature Summary

Table 2-2 lists several of the aforementioned studies, along with their contributions and limitations in scope. In summary, it has been repeatedly shown that contact microphones transmit little background noise. It has also been shown that the transmitted speech from contact microphones in some locations suffers from poor clarity due to the attenuation of high frequency content and the loss of consonant information. However, other papers have shown that contact microphones can nevertheless yield an understandable signal. Reconstruction filters have been shown to help restore attenuated high frequencies to improve transmitted speech clarity.

Table 2-2 Selected literature summary

Study	Subjects	Location(s)	Sounds	Excitation	Data Analysis
Moser & Oyer, 1958	3 Male	16 head/neck	12 Vowels	Voice	Subjective intelligibility, intensity
Snidecor et al., 1959	1 Male	8 head/neck	4 Vowels	Voice	Intelligibility, intensity
Hayes & Meltzer, 1967	1, Gender unreported	Forehead	Sentence	Voice	Intelligibility, signal-to-noise
Norton et al., 1993	Not reported	Neck	1 Vowel	Shaker	Frequency response function
Meltzner et al., 2003	7 Male, 7 Female	3 Neck	3 Vowels	Shaker	Frequency response function
Acker-Mills et al., 2004	9 Male, 1 Female	Throat	Various words	Voice	Spectral analysis, time waveform, intelligibility, signal-to-noise
Horacek et al., 2004	1, Gender unreported	Forehead	3 Vowels, sentences	Voice	Frequency response function spectral analysis, subjective inteligibility
Shimamura et al., 2005	2 Male, 2 Female	Top of head	3 Vowels, sentences	Voice	Frequency response function, spectral, intelligibilty
Svec et al., 2005	10 Male, 7 Female	Jugular notch	Normal speech	Voice	Skin acceleration level / sound pressure level

Lacking in the literature are detailed reports of the frequency response of head and neck skin vibration during speech production at various anatomical locations, as well as the frequency analysis of sounds other than vowels at numerous locations on the head and neck. The objective of this study is to assist in the development of improved contact microphones and associated signal processing methods by contributing to an improved understanding of the frequency response of the skin at various locations on the head and neck during speech.

The research presented in this thesis is intended to provide information to researchers and engineers in order to develop improved methods of communication in high noise environments. Improving intelligibility of contact microphones can be of great benefit to military communications, consumer cell phone use, and speech recognition software applications.

3 Head and Neck Frequency Response Characterization

In this chapter the data collection and analysis methods to obtain the frequency response of the skin during speech production are described. Results are reported for power spectral density (PSD), power spectral density summed difference (PSD_{SD}, defined in Section *3.1.2.1*), spectrogram, perceptual ratings, and signal to noise ratio calculations.

3.1 Methods

3.1.1 Experimental Setup

To measure the frequency response of the skin during speech, accelerometers were attached to 15 locations on the face and neck of 14 male and 10 female subjects using medical-grade double-sided adhesive tape (see Figure 3-1 and Table 3-1). Prior to accelerometer placement the subjects removed oil and/or makeup with an alcohol prep pad to ensure adequate adhesion. All testing was done with Institutional Review Board (IRB) approval and in accordance with IRB policies. Note that different types of transducers, including accelerometers, can be used as contact microphones. Hereafter when specifically discussing the present study, accelerometers are referred to, but the term contact microphones is used in more general discussions.

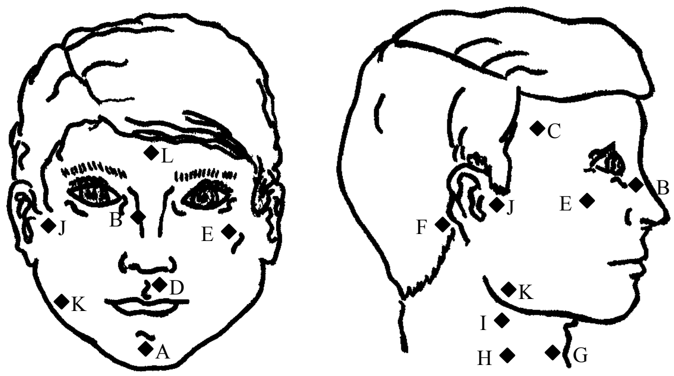

Figure 3-1 Accelerometer placement locations (♦) (image courtesy U.S. Army Research Lab Human Research & Engineering Directorate).

Table 3-1 Accelerometer locations and specifications (locations identified in Figure 3-1).

Location	Accelerometer Model #	Mass [g]	Sensitivity [mV/(m/s^2)]	Frequency range [Hz] (+- 5%)
A - Chin	352A24	0.8	10.2	1 to 8000
B – Nasal Bone	352A56	1.8	10.2	0.5 to 10000
C - Temple	352A24	0.8	10.2	1 to 8000
C* - Temple	333B53	7.5	102	0.5 to 3000
D - Upper Lip	352A24	0.8	10.2	1 to 8000
E - Zygomatic	352A56	1.8	10.2	0.5 to 10000
E* - Zygomatic	333B53	7.5	102	0.5 to 3000
F - Mastoid Process	352A56	1.8	10.2	0.5 to 10000
F* - Mastoid Process	333B53	7.5	102	0.5 to 3000
G - Over Vocal Folds	352A56	1.8	10.2	0.5 to 10000
H - Neck Lateral to G	352A56	1.8	10.2	0.5 to 10000
I - Neck Superior to H	352A56	1.8	10.2	0.5 to 10000
J - Condylar Process	352A24	0.8	10.2	1 to 8000
K - Angle of Mandible	352A56	1.8	10.2	0.5 to 10000
L - Forehead	352A56	1.8	10.2	0.5 to 10000

All accelerometers were manufactured by PCB Piezotronics; model numbers and relevant specifications are listed in Table 3-1. The higher sensitivity accelerometers (C*, E*, and F*), used with the intent of improving signal detection at locations that were previously identified as having low signal-to-noise ratios, were placed on the opposite sides of the head as accelerometers C, E, and F, respectively. Each subject was seated upright in a chair with a head rest to minimize head movement during testing. Due to the larger sizes of accelerometers C*, E*, and F*, their wires were suspended from above the subject's head, with a slight vertical tension applied to roughly compensate for the accelerometers' masses. The wires for all other accelerometers were attached to the head rest to minimize the torque on the skin due to the weight of the wires. The pressure of the accelerometers on the skin was not measured.

A 0.635-cm Larson-Davis 2520 microphone with a Larson-Davis PRM910B preamplifier was placed 45 degrees off-axis, approximately 6.4 cm from the subject's mouth, and was used to simultaneously acquire audible speech data. The microphone and preamplifier were connected to a power supply (Larson-Davis 2221, 20 dB gain). All testing was done in a single-wall Acoustic Systems sound-attenuating booth with a noise floor of approximately 42 dB SPL. The size of the booth was 2.1 m × 2.1 m. The booth had an absorption coefficient of 0.79 at 125 Hz and 1.08 at 250 Hz. (See Fig. A-1 for a image of the sound booth test setup.)

The accelerometer and microphone signals were fed into a National Instruments NI PXI-1042Q data acquisition system with two eight-channel NI PXI-4472 ports (located outside of the booth). All signals were simultaneously sampled at 40 kHz; LabVIEW 7.1 was used for data collection.

The subjects were asked to sit as motionless as possible. A computer screen in the booth displayed instructions to the subjects regarding what to say, including signals for when to start and stop speaking. The subjects sustained the vowels /æ/, /u/, /ɔ/, /i/, the nasals /m/ and /n/, and the fricative /f/ for about 4 to 5 seconds each. The subjects also said, “Rice is often served in round bowls,” a phrase from the phonetically balanced Harvard sentences (IEEE, 1969). To guide vocal effort the subjects were asked to speak in a normal conversational voice. The average sound pressure level over all subjects and sounds was 64.4 dB re 20 μPa @ 6.4 cm with a standard deviation of 5.8 dB. After completing the sounds and phrases in a quiet environment, the subjects were given ear plugs and asked to repeat the same tasks in the presence of 95 dB white noise, which is a sound intensity similar to that of a lawn mower or of a small airplane cockpit. This intensity of white noise was chosen to match the intensity level in the Acker-Mills (2004) study. A separate microphone in the booth allowed the researchers to hear the subjects to ensure proper completion of each speaking task.

3.1.2 Data Analysis

3.1.2.1 Phoneme Analysis

MATLAB was used for signal analysis. Each data set was truncated so that only the portion of the data during which the subject was speaking was analyzed. All signals were also passed through a high-pass filter with a cutoff frequency of 20 Hz to remove low frequency noise from head or jaw motion. Phoneme data were analyzed as follows. For convenience the power spectral density (PSD) was estimated via Welch’s method (Welch, 1967) using the “pwelch” function in MATLAB, with the following parameters: a hamming window with a size of 1024 samples, 50% overlap, and a Fast Fourier Transform (FFT) length of 1024 samples. The accelerometer signals were then normalized to yield the same area under the PSD curve as the microphone signal between zero and five kHz. Five kHz was chosen as the cut-off frequency because higher frequencies are not transmitted in most current communications systems (e.g., telephones). The following equation was used to normalize the accelerometer data:

$$PSD_{i,norm} = PSD_i + \frac{\int_0^{f_c} PSD_{mic}(f)\,df - \int_0^{f_c} PSD_i(f)\,df}{f_c}, \qquad \textbf{(3-1)}$$

where $PSD_{i,norm}$ is the normalized PSD for location i, PSD_{mic} is the PSD of the microphone, PSD_i is the PSD at location i, f is the frequency and f_c is the cutoff frequency (5 kHz). The integrals were calculated using the trapezoidal method.

To compare how well each of the accelerometer signals matched that of the microphone, the absolute value of the difference between the normalized PSD of each accelerometer and that of the microphone signal was found at each frequency and summed from zero to five kHz. This resulted in a single value for each of the accelerometer signals, here referred to as the power spectral density summed difference (PSD_{SD}):

$$PSD_{SD,i} = \sum_{f=0}^{f_c} \left| PSD_{i,norm}(f) - PSD_{mic}(f) \right|, \qquad \textbf{(3-2)}$$

where the $PSD_{SD,i}$ is the power spectral density summed difference of location i. A low PSD_{SD} value indicates little difference between the accelerometer and microphone spectra, and a high PSD_{SD} value indicates little agreement between the accelerometer and microphone spectra. The PSD_{SD} was calculated for each subject, sound, and location, and was then averaged at each location over all subjects to obtain an average PSD_{SD} value for each sound and location.

Each location was given a ranking from 1 to 15 for each subject based on the subject's PSD_{SD}. For example, if location A yielded the lowest PSD_{SD} value for a given subject, the "individual subject rank" for location A for this individual was 1. Additionally, an "average subject rank" was calculated for each location by averaging the individual subject ranks for the corresponding location over all subjects. A rank of 1 indicates the lowest (best) PSD_{SD} value and a rank of 15 indicates the highest (worst) PSD_{SD} value.

3.1.2.2 Phrase Analysis

A spectrogram for the phrase "Rice is often served in round bowls" was calculated for a few selected locations. The spectrogram was generated using the 'spectrogram' function in MATLAB, which calculates the PSD estimate over select time intervals using Welch's method. The same parameters were used to calculate the spectrogram as were used to calculate the PSD as described above. The spectrogram plots were normalized so each accelerometer's spectrogram had the same volume under the surface as the microphone spectrogram. This ensured that each spectrogram had the same dynamic range for comparison.

3.1.2.3 Perceptual Ratings

For four of the subjects (two male and two female), the phrase "Rice is often served in round bowls" was rated by eleven separate individuals (four male, seven female, ages 21 to 30) who listened to recordings from each of the 15 accelerometers and the microphone for all four subjects. All listeners self-reported normal hearing. Fifteen of the recordings were randomly selected and repeated in order to determine intra-rater reliability. The order of these recordings (79 in all) was randomized and each listener heard the recordings in the same order. To ensure that all the recordings had similar volume levels, all signals were passed through a high pass filter with a cutoff frequency of 20 Hz (to remove any low frequency content that could have resulted from jaw or head motion during speech) and normalized.

The voice rating method described by Dromey et. al. (2008) was used to rate each recording's quality. Using a custom MATLAB routine (courtesy Dr. Christopher Dromey), the listeners rated the quality of each recording on a scale from 'bad' to 'good' using a graphical user interface slider bar with continuous values from zero ('bad') to 100 ('good'). The overall quality judgment included factors such as intelligibility, amount of noise, and if the recording was natural-sounding. Intra-rater reliability was found by calculating the Pearson correlation between the first and second rating for the 15 repeated samples for each listener. For the eleven listeners the correlation values ranged from 0.55 to 0.84, indicating that some were inconsistent in their ratings. Therefore, only the five listeners with correlations above 0.73 were included in the study; these five listeners had an average correlation of 0.78. An intraclass correlation coefficient (ICC) was calculated and used to judge interrater reliability. The single measures ICC was 0.717, and the average measures ICC was 0.927, with an *F-ratio* $F(63, 252) = 13.66$ and $p < 0.001$, indicating that the raters were fairly consistent.

As in the PSD_{SD} rankings, the locations were ranked based on the rating from each listener. These rankings were averaged to determine an average rating rank for each location.

3.1.2.4 Signal-To-Noise Ratio

The signal-to-noise ratio (SNR) for the sound /ɔ/ was calculated for each subject and location with and without the presence of background noise. The SNR was calculated using the following equation:

$$SNR = 10\log_{10}\left(\frac{\frac{1}{k}\sum_{n=1}^{k}\left|signal(n)^{2}\right|}{\frac{1}{l}\sum_{m=1}^{l}\left|noise(m)^{2}\right|}\right), \tag{3}$$

where *signal* denotes a portion of the data during which the subject is talking, *noise* is a portion of the data prior to the subject talking, *k* is the number of data points in the signal portion, and *l* is the number of data points in the noise portion. The average SNR was also calculated for each location over all subjects.

3.2 Results

3.2.1 Power Spectral Density

The PSD was used to compare the frequency content of the accelerometer signal to that of the microphone signal. Figure 3-2 shows the normalized PSD for sounds /æ/, /i/, /m/, /u/, and /f/ for one male speaker. Figures 3-2(a,b,d) show that for vowels, the upper lip spectrum matched the microphone spectrum very well up to about 3.5 kHz, while the nasal bone spectrum match was limited to 2 or 3 kHz. For the vowels /æ/ and /i/, the location over the vocal folds followed the general trend of the microphone, but agreement was worse than for the other two locations. For the sound /æ/ the nasal bone spectrum shows a large peak around 2.5 kHz. For the nasal sound /m/, the nasal bone spectrum corresponded well with the microphone up to 3.5 kHz, albeit with large divergence in the 1.5 to 2.5 kHz region. The upper lip and vocal fold signals did not correspond very well to the microphone spectrum for the sound /m/.

Figure 3-2 also shows that the accelerometer spectra are generally attenuated beyond about 4 kHz, although peaks in the spectrum between 4-5 kHz were detected at the location above the upper lip (Fig. 3-2(b-c)). Figure 3-2(e) shows that for /f/ there is only slight agreement between the accelerometer spectrum and the microphone spectrum.

3.2.2 Power Spectral Density Summed Difference

Figures 3-3 and 3-4 and Tables 3-2 and 3-3 show the average PSD_{SD} over 0-5 kHz for all male and female subjects. The tables are sorted according to the average PSD_{SD}. For visual clarity only one side of the standard deviation bars are displayed. The order of locations along the *x*-axis for these figures is according to the location's average ranking (Sec. 3.2.3). These figures show that the overall trends in PSD_{SD} vs. location are similar between the male and female speakers, although they are generally slightly lower for the female speakers. They also indicate that one particular location does not always have the lowest PSD_{SD} for each sound or gender. These tables and figures show that for vowels, the upper lip signal matched the spectrum of the microphone signal better than the other locations, while for nasals, the nasal bone signal had a spectrum that matched that of the microphone better than the other locations; this was consistent for both male and female speakers. The standard deviation of the male and female average PSD_{SD} values is comparable, with the male values being slightly higher for most sounds and locations.

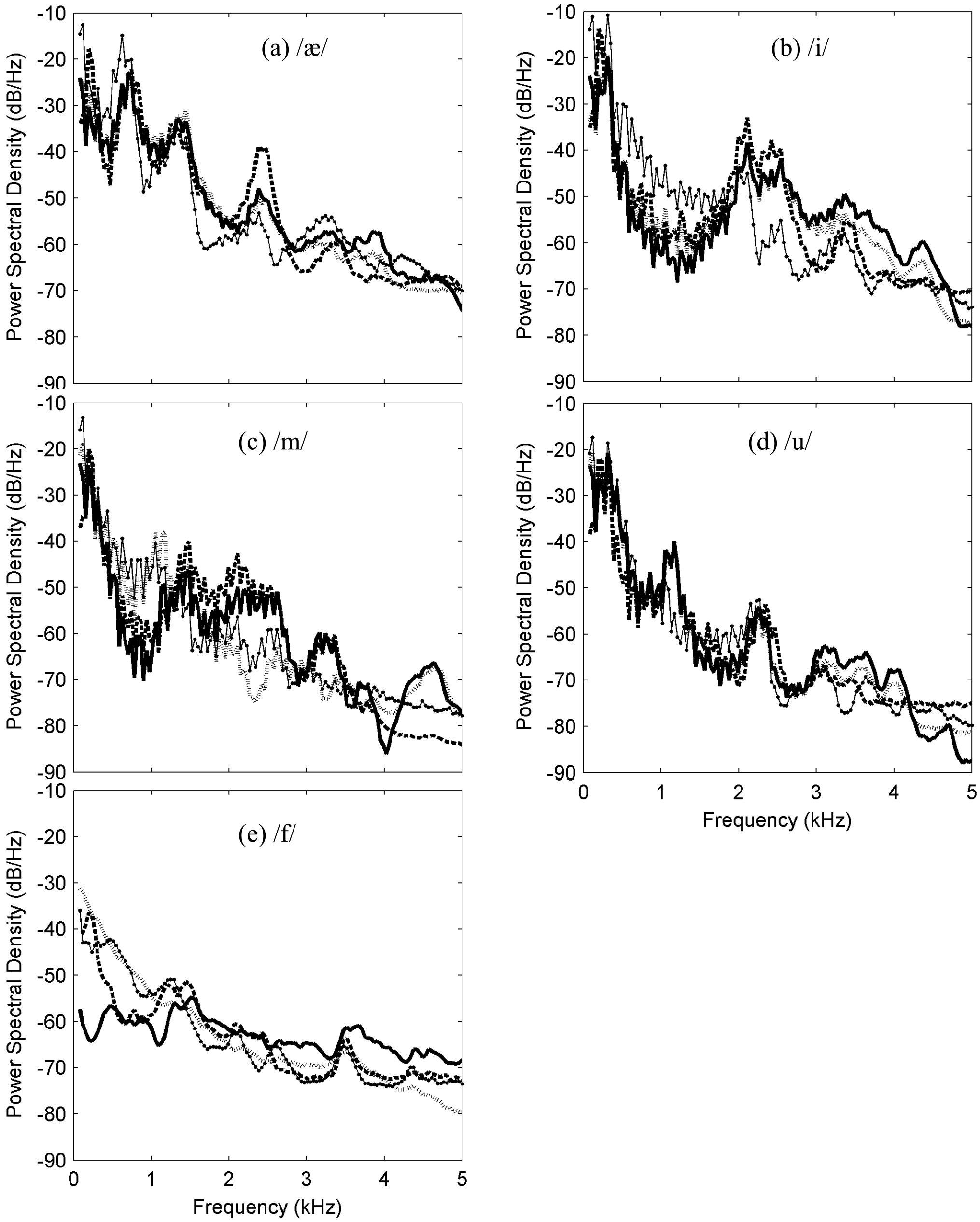

Figure 3-2 PSD for one male subject; —— Microphone; ······ Above upper lip; ------- Nasal bone; —•— Over the vocal folds. a) /æ/; b) /i/; c) /m/; d) /u/; e) /f/.

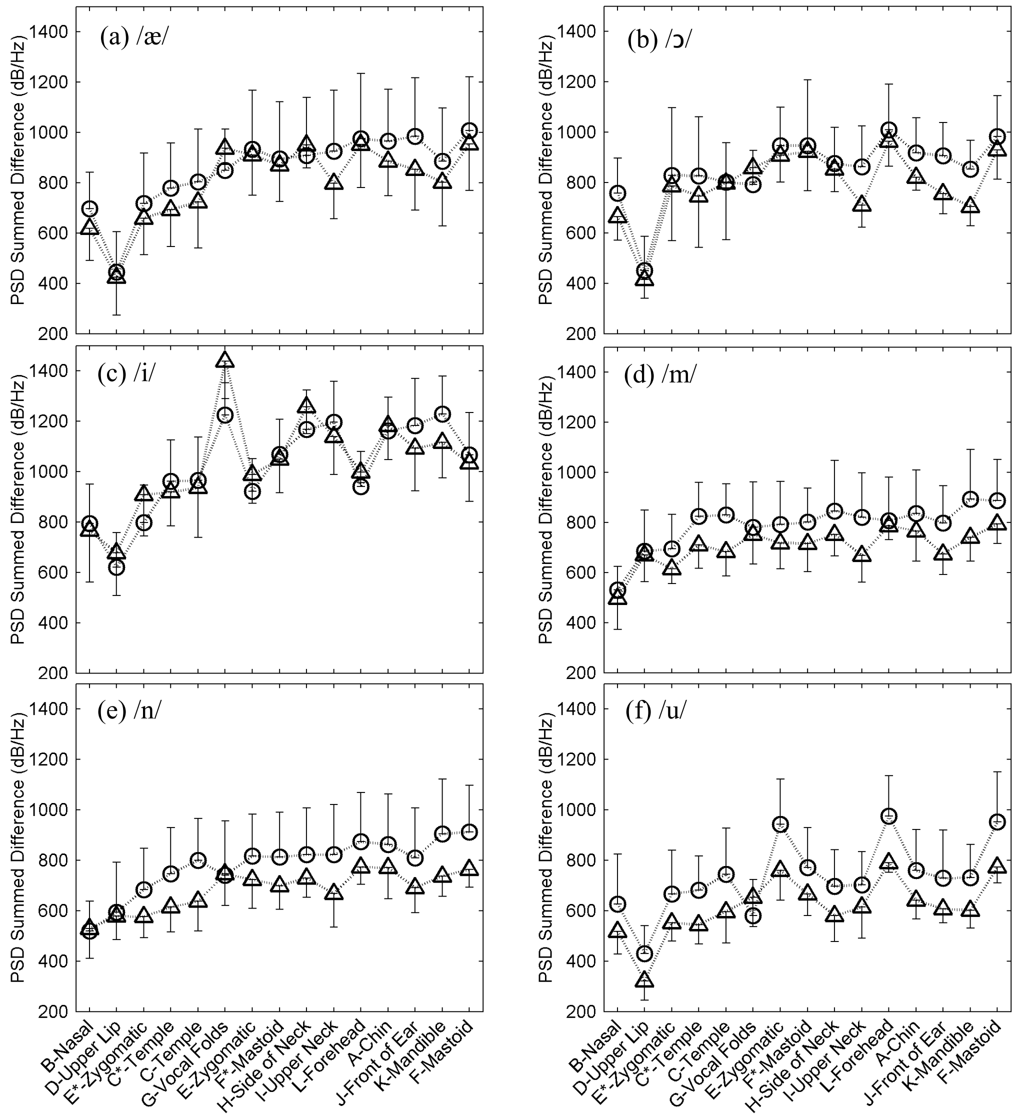

Figure 3-3 Normalized PSD_{SD} over 0-5 kHz. ○: Male speakers; ∆: Female speakers. a) /æ/; b) /ɔ/; c) /i/; d) /m/; e) /n/; f) /u/.

3.2.3 Power Spectral Density Ranking

Tables 3-4 and 3-5 show the male and female average PSD_{SD} ranks, respectively. These tables are sorted according to the average rank over all sounds for each location. The tables also show the average rank for each sound at each location. These tables show that for the seven sounds analyzed, the top five ranked locations are the same for both male and female speakers. These locations are the nasal bone, above the upper lip, the zygomatic*, and the two temple locations.

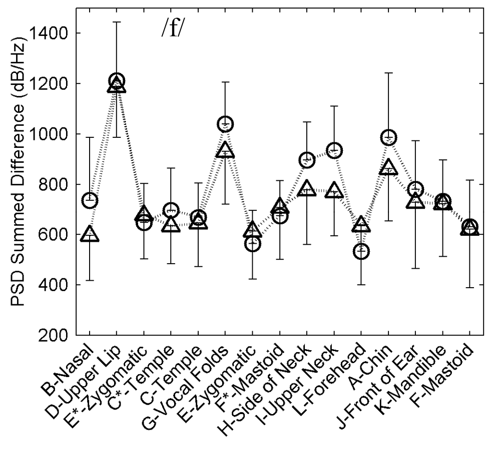

Figure 3-4 Sound /f/ normalized PSD_{SD}. ○: Male speakers; Δ: Female speakers.

Table 3-2 Average PSD_{SD} male speakers, all values dB/Hz.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
D	**445**	**451**	**621**	**431**	684	594	1212	634
B	696	758	794	627	**531**	**520**	737	666
E*	717	829	798	667	694	684	646	719
C*	779	827	961	681	824	747	697	788
C	804	801	965	745	830	799	668	802
E	933	947	922	943	791	816	563	845
F*	894	946	1069	771	801	812	674	853
G	849	792	1225	581	780	740	1040	858
L	975	1009	941	975	807	874	**533**	873
J	985	907	1183	729	798	808	781	884
H	908	876	1167	697	846	822	896	887
K	886	852	1228	730	892	905	733	889
I	925	863	1196	703	821	822	934	895
F	1008	982	1067	952	886	912	630	920
A	964	918	1162	759	835	863	985	927

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 3-3 Average PSD_{SD} female speakers, all values dB/Hz.

Location	Sounds /æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	Average
B	618	663	767	517	**497**	**530**	**595**	598
D	**424**	**414**	**677**	**321**	669	578	1189	610
E*	659	786	908	551	615	576	678	682
C*	693	746	919	545	709	615	634	694
C	724	798	937	596	683	637	645	717
I	797	710	1138	614	669	668	772	767
J	852	756	1093	607	674	692	730	772
K	802	705	1115	602	740	737	724	775
E	911	907	988	759	718	723	613	803
F*	870	923	1048	666	716	699	707	804
F	955	929	1033	772	794	763	620	838
L	953	964	998	790	786	772	634	843
H	950	854	1256	580	753	730	779	843
A	886	820	1183	642	765	770	863	847
G	936	859	1439	654	752	746	932	902

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 3-4 Average PSD_{SD} subject rank, male speakers.

Location	Sounds /æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	Average
B	3.6	5.1	3.6	4.7	**1.4**	**1.7**	7.3	3.9
D	**1.1**	**1.2**	**1.4**	**1.6**	5.4	3.2	14.2	4.0
E*	3.9	5.8	2.9	5.6	3.2	4.4	5.3	4.4
C*	5.1	5.6	5.5	6.1	9.1	5.9	7.1	6.3
C	5.9	6.0	5.9	8.2	10.0	8.6	5.6	7.2
G	7.2	5.9	12.3	3.8	6.9	5.9	13.4	7.9
E	9.9	11.1	5.0	13.5	7.3	8.7	2.9	8.3
F*	8.6	9.3	8.6	9.6	8.3	8.4	5.9	8.4
H	9.0	8.3	11.4	6.7	9.3	8.7	11.2	9.2
L	11.1	12.9	6.0	13.4	8.9	10.6	**1.9**	9.3
I	9.9	8.1	12.2	7.4	8.7	8.9	12.1	9.6
J	12.1	9.9	12.1	8.0	8.7	9.0	8.6	9.8
K	8.9	7.8	13.2	8.6	11.1	12.9	7.7	10.0
A	11.1	10.8	11.0	8.9	9.5	10.6	11.9	10.5
F	12.7	12.4	8.8	13.9	12.2	12.5	5.0	11.1

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 3-5 Average PSD_{SD} subject rank, female speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
B	2.6	4.4	2.6	4.8	**2.4**	**3.2**	**3.5**	3.4
D	**1**	**1**	**1.5**	**1**	6.8	3.9	14.6	4.3
E*	3.6	6.9	5	4.9	3.2	3.3	6.7	4.8
C*	4.4	5.6	5	4.8	8.4	4.5	6	5.5
C	5.1	7.2	5.4	7.3	6.8	6	5.8	6.2
I	6.9	5	10.6	7.1	5.8	6.7	9.7	7.4
J	8.6	6.7	9.8	8	5.5	7.9	8.3	7.8
K	7.6	4.7	10.2	7.8	9.6	10.5	8.2	8.4
E	10.4	11.6	6.3	12	8	9.5	4.7	8.9
F*	10.1	11.6	8.2	9.9	8.4	8.5	8.1	9.3
H	12.2	10.4	13.4	6.2	10.1	10.1	9.5	10.3
A	10.4	8.7	11.7	9.4	11.4	11.3	11.9	10.7
L	12.3	13.4	7.6	14	11.1	11.6	5.3	10.8
F	13.1	12.4	7.7	13.5	12.4	12.1	5.3	10.9
G	11.7	10.4	15	9.3	10.1	10.9	12.4	11.4

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

For one female subject saying the phrase "Rice is often served in round bowls," each location was rated by the author on a scale of good, fair, poor and very poor quality in order to have a general idea of how PSD_{SD} values corresponded to signal quality. For this subject, PSD_{SD} values under 500 dB/Hz corresponded to signals of good and fair quality. PSD_{SD} values ranging from 500 dB/Hz to 700 dB/Hz corresponded to signals of fair and poor quality. PSD_{SD} values over 700 dB/Hz corresponded to signals of very poor quality. It should be noted that these subjective ratings are for one subject only and may not correspond directly to the entire population. Further studies should be done to identify which ranges of PSD_{SD} correspond to speech signals of adequate quality.

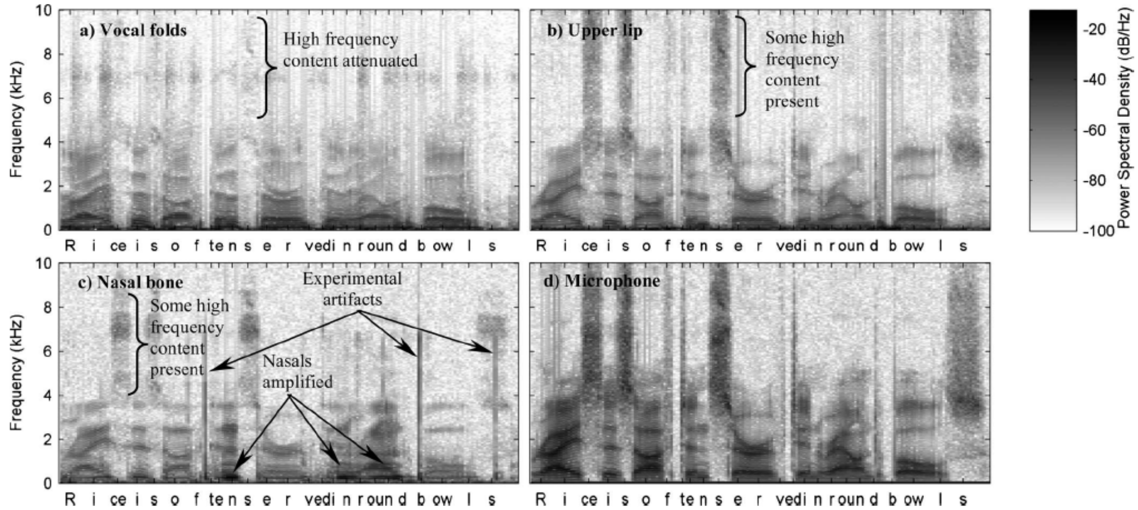

Figure 3-5 Spectrogram of a male subject saying "Rice is often served in round bowls." a) Over the vocal folds; b) above the upper lip; c) over the nasal bone; d) microphone. Experimental artifacts were only present in the nasal bone recording (e.g., noise caused by wire motion and tapping) and are noted.

3.2.4 Spectrogram

Figure 3-5 shows spectrograms for the locations over the vocal folds, above the upper lip, on the nasal bone, and for the microphone for a male subject saying, "Rice is often served in round bowls." Figure 3-5d shows that for nasals and vowels, most of the frequency content of interest is below 5 kHz. However, for the fricatives /f/ and /s/ there is much higher frequency content. Figure 3-5a shows that the location over the vocal folds attenuates most of the fricative high frequency content above 4 kHz. This is in general agreement with the Dupont et al. (2004) study which found that frequencies above 3 kHz were attenuated by a throat microphone worn near the thyroid cartilage. Figures 3-5b and 3-5c show that the upper lip and nasal bone are able tosense some of the high frequency content of the fricatives. Figure 3-5c also shows that the nasal sound /n/ is disproportionately amplified at low frequencies; this amplitude increase was found to be both noticeable and unnatural-sounding when listening to recordings of the nasal bone signals.

3.2.5 Perceptual Ranking

Figure 3-6 shows a comparison of the average perceptual rating rank and the average PSD_{SD} rank for the four subjects included in the perceptual study. The dashed line has a slope of one, indicating where the data would lie given a direct one-to-one correlation between the two ranking methodologies. The data show that, for the most part, a higher rating rank corresponds to higher PSD_{SD} rank. However, there were three outliers: the zygomatic (E), mastoid (F), and upper neck (I). The upper neck has a much lower perceptual rank than PSD_{SD} rank while the zygomatic and mastoid have perceptual ranks higher than the PSD_{SD} rank. The correlation coefficient of the average rating rank to the average PSD_{SD} rank is 0.57 for all locations. With the three outliers removed, the correlation coefficient improves to 0.90.

3.2.6 Signal-To-Noise Ratio and Background Noise

Figure 3-7 and Table 3-6 shows the average SNR for the sound /ɔ/ over all subjects with and without the presence of 95 dB white noise in the background. These data show that all locations are affected somewhat by the presence of background noise, some considerably more than others. The areas least affected are the neck locations (G, H, I), the angle of the mandible (K), and the mastoid (F). Although reduced significantly in the presence of noise (~10 dB), the nasal bone SNR is still in the top four, in terms of the no-noise SNR, with the locations on the neck.

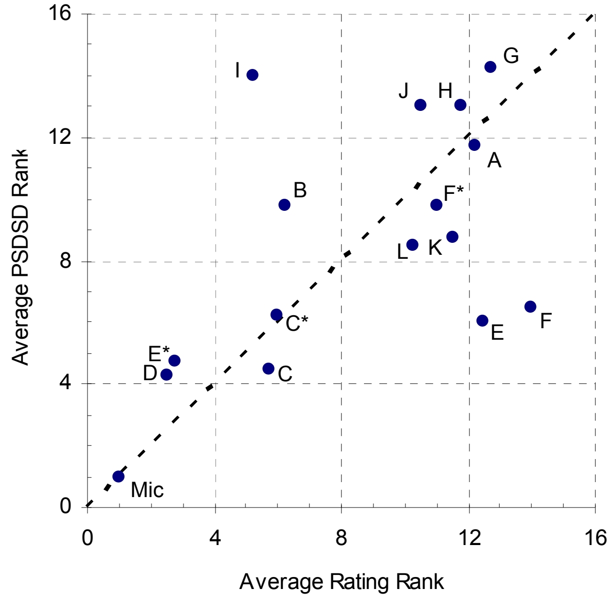

Figure 3-6 Comparison of PSD_{SD} ranking and perceptual rating ranking for four randomly selected subjects.

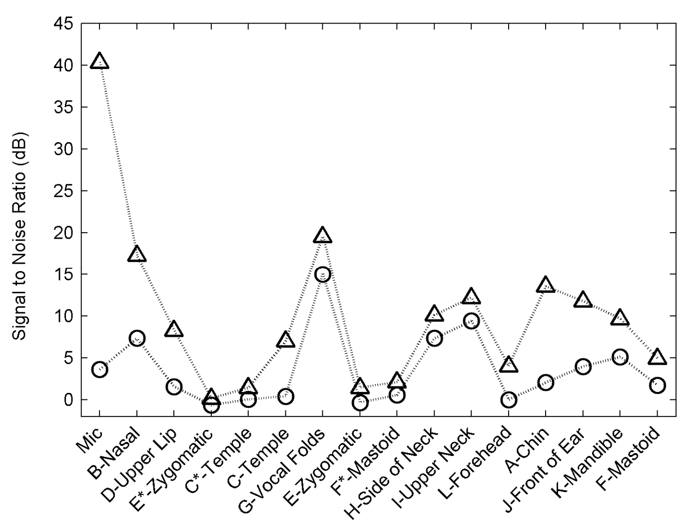

Figure 3-7 Sound /ɔ/ average SNR over all subjects. ∆: No background noise; ○: 95 dB background noise.

Table 3-6 Sound /ɔ/ average SNR over all subjects.

Location	SNR (dB)	
	No Noise	95 dB Noise
Microphone	40.3	3.8
G- Vocal Folds	19.4	15.0
B-Nasal Bone	17.2	7.9
A-Chin	14.6	2.4
I-Upper Neck	12.7	9.3
J-Front of Ear	11.4	3.7
H-Side of Neck	10.5	7.5
K-Mandible	9.7	5.2
D-Upper Lip	8.4	1.6
C-Temple	7.1	0.5
F-Mastoid	5.4	1.9
L-Forehead	4.4	0.1
F*-Mastoid	3.1	0.7
C*-Temple	1.8	0.2
E-Zygomatic	1.6	-0.2
E*-Zygomatic	0.49	-0.4

3.3 Discussion of Results

3.3.1 Influence of Location

Figure 3-3 shows that after about the best three to five locations, the PSD_{SD} values for the remaining locations remain comparable. This indicates that these first few locations may yield signals that match the microphone signal better than the rest. The signal spectra from the remaining locations, however, should all match the spectrum of the microphone similarly.

No single location was found to have the lowest PSD_{SD} for all sounds; that is, some locations had lower PSD_{SD} values for some sounds than others. For both male and female speakers, the vowels had the lowest PSD_{SD} values above the upper lip. Vowels are shaped in the mouth using the lips and tongue and are transmitted primarily through the open mouth. The accelerometer closest to the area where the sounds are shaped and transmitted was located above the upper lip. It is, however, somewhat surprising that the upper lip had a low PSD_{SD} value given its compliant tissue composition. This could be attributed to the accelerometer possibly picking up the speech signal through the air as opposed to primarily through the skin. Table 3.6 shows that the upper lip SNR drops from 8.4 to 1.7 dB in the presence of 95 dB background noise. This suggests a potentially rather high sensitivity of the accelerometer on the upper lip to airborne acoustic waves from the talker's speech.

The nasal sounds /m/ and /n/ are also influenced by mouth shape, but no air flows through the mouth and thus the sound is mainly transmitted through the nasal passages. This resulted in the accelerometer on the nasal bone having lower PSD_{SD} values for the nasal sounds than at other locations.

A low PSD_{SD} value indicates good agreement between the spectra of the accelerometer and the microphone signals, and this usually – but not always – correlates with a clearer and/or higher perceived quality accelerometer signal. This is discussed more in Section 3.3.5.

During the study the same type of transducer was used at each location, and no attempt was made to improve sound transmission via impedance matching. The impedance of areas above soft tissue, such as on the neck, may be very different from the impedance of tissue over bony structures. This is a limitation of the current study and it is recommended that further research investigate the implementation of devices that match the impedance of the various sites.

One possible method of identifying ideal contact microphone locations is to use perturbation theory to identify locations of formant pressure anti-nodes along the vocal tract. There may be higher skin acceleration levels at the skin surface near anti-nodes. Perturbation theory could also yield insight as to why some locations performed better than others. However, perturbation theory is only applicable within the vocal tract itself, and doesn't extend to the skin surface; thus, further study would be needed to identify a relationship between formant pressure anti-node locations and quality of skin vibration signals near these locations. Also, these locations can change as the articulators move – i.e., they will be phoneme-specific.

3.3.2 Difference between Male and Female Speakers

Tables 3-2 and 3-3 show that when using the PSD_{SD} to compare the spectra of different locations, the top five locations are the same for both male and female speakers, but the other locations have different rankings between gender. One notable difference is that for males, the location over the vocal folds (G) had the sixth highest average rank, while for females it ranked last. This suggests that throat microphones placed near the vocal folds may be expected to match the microphone spectrum poorly for female speakers. However, for female speakers, the location on the upper neck (I) had the sixth best overall rank, indicating that if the neck is the preferred or necessary location for a contact microphone, the placement on the neck may benefit from vertical adjustment based on the speaker's gender.

Tables 3-2 and 3-3 show that females generally had lower PSD_{SD} values than males, indicating that, on average, the female speakers' accelerometer signals matched the microphone signals better than those from the male speakers. This result is generally consistent across locations and subjects, but the differences between the male and female PSD_{SD} average values were rather small. There were a few exceptions to this, however. Male speakers on average had lower PSD_{SD} values than females over the vocal folds for sounds /u/, /æ/, /n/, /ɔ/ and /i/. For the sound /i/, males had a lower PSD_{SD} for the locations above the upper lip, zygomatic, side of neck, forehead and chin.

The average fundamental frequency for male speakers over the vowel and nasal sounds was 116.8 Hz with a standard deviation of 11.1 Hz. The average fundamental frequency for female speakers over the vowel and nasal sounds was 221.7 Hz with a standard deviation of 35.1 Hz. It is thus possible that some of the gender-associated differences in these results could simply be attributed to the male speakers having lower fundamental frequencies than the female speakers, and may not necessarily be due to general differences between genders in terms of superficial anatomical composition.

3.3.3 Influence of Accelerometer Sensitivity

Accelerometers were placed on both sides of the head at the temple, zygomatic bone, and the mastoid process. An accelerometer with 10.2 mV/(m/s^2) sensitivity was placed on one side while an accelerometer with 102 mV/(m/s^2) sensitivity was placed on the other at each of these three locations. For the present discussion, it is assumed that the speech signals were transmitted symmetrically to each side of the head. It can be seen from Fig. 3-3 and Tables 3-2 to 3-6 that accelerometers with the higher sensitivity (indicated by *) in most cases yielded lower PSD_{SD} values and in all cases yielded better *overall* rankings than the corresponding lower sensitivity accelerometers. However, it is interesting to observe from Tables 3-4 and 3-5 that the higher sensitivity accelerometers all yielded poorer rankings for /f/. This is opposite to what was expected, and is attributed to the following. The lower-sensitivity accelerometers in these locations have very low output in response to the low amplitude /f/ vibrations; thus the spectra for these accelerometers are essentially due to noise. At the same time, the microphone spectra for /f/ were relatively flat. Therefore, when normalized, the PSD plots of the lower sensitivity devices appear to show a frequency response that matches the microphone better than the higher sensitivity devices. Notwithstanding this difference in fricative response, the *overall* lower ranking indicates that a higher sensitivity accelerometer (or other type of contact microphone) could be used to transmit a signal that better matches the spectrum of a clean microphone signal, possibly increasing the overall speech clarity. However, this improvement may only be marginal. In some cases (e.g., C* and C for male and female speakers) the difference between the higher and lower sensitivity accelerometers was very small, which indicates a higher sensitivity device may not yield a substantial improvement at all locations.

Figure 3-7 shows that the higher sensitivity accelerometers yielded *lower* SNR than the lower sensitivity accelerometers placed at symmetric locations on the head. However, even though the SNR was lower for higher sensitivity accelerometers in these locations, Fig. 3-7 shows their perceptual rating to be better than the lower sensitivity accelerometers.

Figures 3-6a-c show that the various vibration transmission paths clearly attenuate the intensity of higher frequencies, but they are in some places still detectable. The higher sensitivity accelerometers have signals that better match the frequency content of the microphone signal and are preferred by listeners over lower sensitivity accelerometers. For real world applications signal processing could be used to amplify the attenuated high frequency content in order to increase speech clarity. However, size and cost are important considerations for practical implementation, and higher sensitivity devices often cost more, are typically larger, and the benefit may not be substantial enough to outweigh these disadvantages.

3.3.4 Fricative Sound Transmission and Formation

The fricative /f/ is formed by unsteady airflow generated at a small opening between the teeth and lower lip; the vocal folds are not vibrating. Consequently /f/ is relatively quiet. It was speculated that due to close proximity to the sound source, the upper lip would possibly yield the best ranking for /f/. However, for male and female speakers, the upper lip yielded the worst ranking; the best ranked location for males for /f/ was on the forehead. This ranking result and consideration of Fig. 3-2(e) suggest that fricative sounds may be too highly attenuated by body tissue to be adequately detected at most locations on the skin. This conclusion is further supported by noting that when we compared the PSD of a normalized white noise signal to the microphone PSD for /f/, the white noise ranked better than all other locations. Although the data collection and analysis methods seem to work well for the voiced sounds, the results for the fricative sound /f/ indicate that these methods may not work well for unvoiced sounds. Further investigation into the transmission to the skin surface of fricatives and other phonemes with significant non-voiced acoustic sources is necessary to better understand this behavior.

3.3.5 Power Spectral Density vs. Perceptual Ratings as a Metric

Although some of the high frequency content is attenuated by the body tissue, the speech signals that were detected on the skin surface were still mostly understandable. Figure 3-6 shows that the average perceptual rating rank generally increases with average PSD_{SD} rank. The major outliers to this are locations E, F, and I. The recordings at locations E and F had some experimental noise from wires tapping each other due to jaw motion during speech that overpowered the speech signal. This may have lowered the perceived quality of the recording for these locations. The recordings for location I (upper neck) have very large SNRs, which is attributed to the large vibration amplitudes due to close proximity to the vocal folds. Since the signal at the upper neck lacks high frequency content (which worsens the PSD_{SD} rank), the signal is slightly “muffled”; however it is also relatively devoid of background noise, possibly resulting in a higher perceived quality. This indicates that while some locations may not yield spectra that match the microphone spectrum well, they still can yield signals that are understandable and of reasonable perceived quality. The correlation coefficient jumps from 0.57 with all locations to 0.90 without the outlier locations E, F and I which indicates that for most locations there is a relationship between PSD_{SD} rank and rating rank. In other words, for most locations the degree to which the accelerometer’s spectrum matches the microphone’s spectrum is relatively well-correlated with the perceived quality of the signal. The correlation is not “fail-safe,” though, as the outliers show.

It may seem surprising that the forehead location (L) did not rate better, given that it is one of the preferred locations for commercially available contact microphones. It is important to note that the raters in this study were not just judging intelligibility, but rather were asked to rate quality. The recordings at the forehead, while understandable, had low signal amplitudes which resulted in a lower SNR than some of the locations that rated better. An average rating rank around 10 for the forehead location does not imply poor intelligibility; rather it just indicates that the recordings that were ranked lower were preferred by the raters. A higher sensitivity accelerometer located at the forehead may have resulted in ratings that favored this location.

Another notable result is that both the PSD_{SD} analysis and the perceptual ratings indicate that the location over the vocal folds (G) was one on the worst ranked locations. This is interesting because the area near the vocal folds is one of the most commonly used locations for contact microphone placement, most likely due to the high SNR at this location. These results indicate that in order to have a signal that is preferred by listeners, this region is not a good location for contact microphone placement.

It is also interesting to note that there seem to be three major groupings in the perceptual rankings. The upper lip (D) and the zygomatic (E*) have average perceptual rating ranks between 2 and 3. The upper neck (I), nasal bone (B) and the temple (C and C*) also have similar perceptual rating ranks between 5 and 6. The remaining locations have average perceptual ranks between 10 and 14.

3.3.6 Signal-To-Noise Ratio

Previous studies have emphasized the superior signal-to-noise ratios (SNRs) of contact microphones in high noise environments. For the sound /ɔ/ using accelerometers, it was found that this is the case for some locations, although other locations yielded generally poor SNRs. For the sound /ɔ/ Fig. 3-7 shows that in quiet environments, all accelerometer signals had a SNR much lower than the microphone signal. However, with 95 dB background noise, five locations had better SNRs than the microphone. These locations are the neck (G, H, and I), the nasal bone (B), and the angle of mandible (K). An important consideration to note is that the accelerometers used in this study were not directional or shielded from external acoustical energy. The drop in SNR seen in many of the locations may be reduced by implementing an accelerometer with its external surface acoustically shielded from the surrounding environment. Further, improved impedance matching and application of an external force on the accelerometer may improve noise elimination. These results are for only one vowel sound but nevertheless show that not all skin locations yield improved SNR. Chapter 4 reports the effects of location on SNR for other sounds.

3.3.7 Implications for Contact Microphone Performance

The results of this study suggest that a multi-location device may be useful due to different phonemes being sensed better at different locations. An example of this can be seen using Fig. 3-5. The location over the vocal folds senses the low frequency content very well, while attenuating the high frequency content. The nasal bone, however, is able to detect the high frequency content. A combination of the low frequency content from the vocal folds and the high frequency content from the nasal bone could result in a signal that more accurately represents the microphone signal. However, although a given location may transmit speech well, such as above the upper lip, being able to wear the microphone at that location may not be feasible. Therefore, a combination of locations that each individually yield overall lower performance, but that exhibit favorable individual performance for select phonemes, may yield an improved signal when combined.

The present data further elucidate the complications arising from different speech sound production mechanisms, resulting in unevenness of the transmission of the various speech signals through the skin. For example, the nasal bone was found to be highly sensitive to nasal sounds. The consequence of this was that the overall sound level rose greatly during nasal sound production, resulting in unnatural sounding speech.

The contact pressure of the accelerometers on the skin was not controlled or measured, thus the results presented may be influenced somewhat by the accelerometer attachment method used. This illustrates a complication that may arise in real world implementation of contact microphones, as transducers will likely be attached to the skin without perfect control of contact pressures. Further investigation on the effects of varying contact pressure of contact microphones is recommended.

3.4 Summary and Conclusions

The objective of this study was to assist in the development of improved contact microphones and associated signal processing methods by contributing to an improved understanding of the frequency response of the skin at various locations on the head and neck during speech. The speech signal detected on the skin was characterized by attaching accelerometers to the skin and having subjects produce a phrase and various phonemes. The accelerometer and microphone PSDs were compared and the accelerometer locations were ranked according to how well the PSD of the signal corresponded to the PSD of the microphone signal for each phoneme. A spectrogram of the phrase was also used to compare the amplitudes of various frequencies at selected locations. A perceptual listening test was conducted to compare the PSD rankings to perceptual rankings. The SNR was found for each location with and without the presence of background noise to determine sensitivity to external noise. The following conclusions were reached as a result of this study:

- Locations other than on the throat can adequately sense the speech signal from the skin. Some of these locations were found to yield signals with spectra better matching that of the microphone and that yielded higher perceptual ratings than the throat signal.
- While the throat attenuates the signal intensity of high frequencies, some high frequency content can be detected at other locations on the head. Fricative sounds are not transmitted well through soft tissue and should be studied further.
- More sensitive devices may yield a more understandable signal, but practical application may be difficult.
- For the subjects tested, vowels matched the microphone spectra best above the upper lip, whereas nasals matched the microphone spectra best on the nasal bone.
- Using PSD_{SD} as the metric, the locations that best matched the frequency content of the microphone are generally common to both males and females. These locations are the nasal bone, above the upper lip, both temple locations, and the zygomatic*.
- Using perceptual ratings as a metric, the highest rated locations are the nasal bone and the zygomatic* followed by both temple locations, above the upper lip, and the upper neck.

- Perceptual rankings generally follow the PSD_{SD} ranking, with a few outliers. This indicates that locations that match the spectrum of the microphone are also generally (but not always) preferred by listeners.
- With the addition of background noise, for the vowel sound /ɔ/, not all accelerometer placement locations yield improved SNRs. Some locations were found to be sensitive to external noise while others were relatively insensitive to noise.
- Combining signals from multiple locations may be beneficial to achieve high-quality speech transmission.

4 Frequency Response Characterization with External Noise

In this chapter the data collection and analysis methods to obtain the frequency response of the skin during speech production in the presence of background noise are described. Results are reported for the power spectral density summed difference (PSD_{SD}) and PSD_{SD} ranks. Results are also presented for the frequency response of the skin in the presence of background noise without phonation.

4.1 Methods

4.1.1 Experimental Setup

The same locations and procedures where used to collect the accelerometer and microphone signals as in Chapter 3. After completing the sounds and phrases in a quiet environment, as outlined in Section 3.1.1, the subjects were given ear plugs and asked to repeat the same sounds and phrases in the presence of 95 dB white noise. A KRK Systems RP-6 studio monitor was placed 18 inches in front of the microphone and was used to play the white noise such that the measured sound pressure level at the microphone was 95 dB. The studio monitor had a reported frequency response of +/- 1.5 dB from 49 Hz to 20 kHz. The white noise signal spectrum as detected by the microphone is shown in Figs. 4-3 and 4-4. Subjects 2, 6, 9 and 23 had missing data for one or more of the accelerometers for the background noise portion of the testing so they were excluded from the following analysis.

To determine the frequency response of the skin when exposed to external noise without phonation, the subjects where asked to sit still with their mouth and nose alternately open and closed in four configurations while the white noise was played. The four configurations were mouth closed/nose open (CO); mouth and nose open (OO); mouth open/nose closed (OC); and mouth and nose closed (CC). For the "nose closed" configurations the subjects plugged their nose with their thumb and pointer finger, being careful not to disrupt the accelerometers.

4.1.2 Data Analysis During Phonation

4.1.2.1 Power Spectral Density Summed Difference

The power spectral density summed difference was calculated as outlined in Section *3.1.2.1* for each phoneme. However, the noisy accelerometers' PSD was not compared to the noisy microphone's PSD; instead it was compared to the microphone signal recorded earlier in a quiet environment. This was done in order to find out which accelerometer locations are least sensitive to external noise by producing a signal that is most like a clean microphone signal even in the presence of noise. The compared recordings are from the same person making the same sound so the frequency spectrum of the speech should be similar. Nevertheless, it is acknowledged that this is a limitation of this analysis since the clean microphone signal was recorded at a different time than the noisy accelerometer signals. Another consideration is that with earplugs and in the presence of noise, the subjects may have spoken differently due to the Lombard effect, which is the tendency for people to speak louder in the presence of background noise. If the subjects raised their voice, the sound source spectrum changes (upper harmonics become stronger) and articulatory movements also increase in displacement and velocity.

4.1.2.2 Power Spectral Density Summed Difference Ranking

Each location and the noisy microphone were given a ranking from 1 to 16 for each subject based on the subject's PSD_{SD}. For example, if location A yielded the lowest PSD_{SD} value for a given subject, the "individual subject rank" for location A for this individual was 1. Additionally, an "average subject rank" was calculated for each location by averaging the individual subject ranks at the corresponding location over all subjects. A rank of 1 indicates the lowest (best) PSD_{SD} value and a rank of 15 indicates the highest (worst) PSD_{SD} value.

4.1.3 Data Analysis Without Phonation

With the 95 dB white noise playing and no phonation, the PSD was estimated for each mouth and nose configuration described in Section 4.1.1 for each location for one male subject to show an example of the variation from the average. The PSD was also averaged over all subjects to obtain an average PSD for each location. The PSD was estimated via Welch's method (Welch, 1967) using the "pwelch" function in MATLAB, with the following parameters: a hamming window with a size of 1024 samples, 50% overlap, and a Fast Fourier Transform (FFT) length of 1024 samples. For both the individual subject and the average over all subjects, the PSD of a portion of the signal recorded prior to white noise onset was estimated to find the noise floor of the microphone and each accelerometer. For each mouth/nose open/closed configuration the noise floor was comparable. For clearer presentation in the figures the average noise floor for the microphone and each location is presented.

4.2 Results

4.2.1 Frequency Response with External Noise During Phonation

4.2.1.1 Power Spectral Density Summed Difference

Figures 4.1 and 4.2 show the average PSD_{SD} over 0-5 kHz for all male and female subjects. The data are sorted according to the overall average PSD_{SD} for the male subjects. A low PSD_{SD} value indicates better agreement between the microphone and the accelerometer. For visual clarity in the figures only one side of the standard deviation bars are displayed (males on top, females on bottom). Similar to the results found in Section 3.2.2, these figures show that the overall trends in PSD_{SD} vs. location are similar between the male and female speakers. These figures also indicate that one particular location does not always have the lowest PSD_{SD} for each sound or gender. Unlike for the clean signals in Chapter 3, one location did not have the lowest PSD_{SD} for all the vowels or all the nasals. (Tables A-2 and A-3 in Appendix A contain the average background noise PSD_{SD} for male and female speakers.) These figures also show that most of the noisy accelerometer location signals matched the spectra of the clean microphone better than that of the noisy microphone.

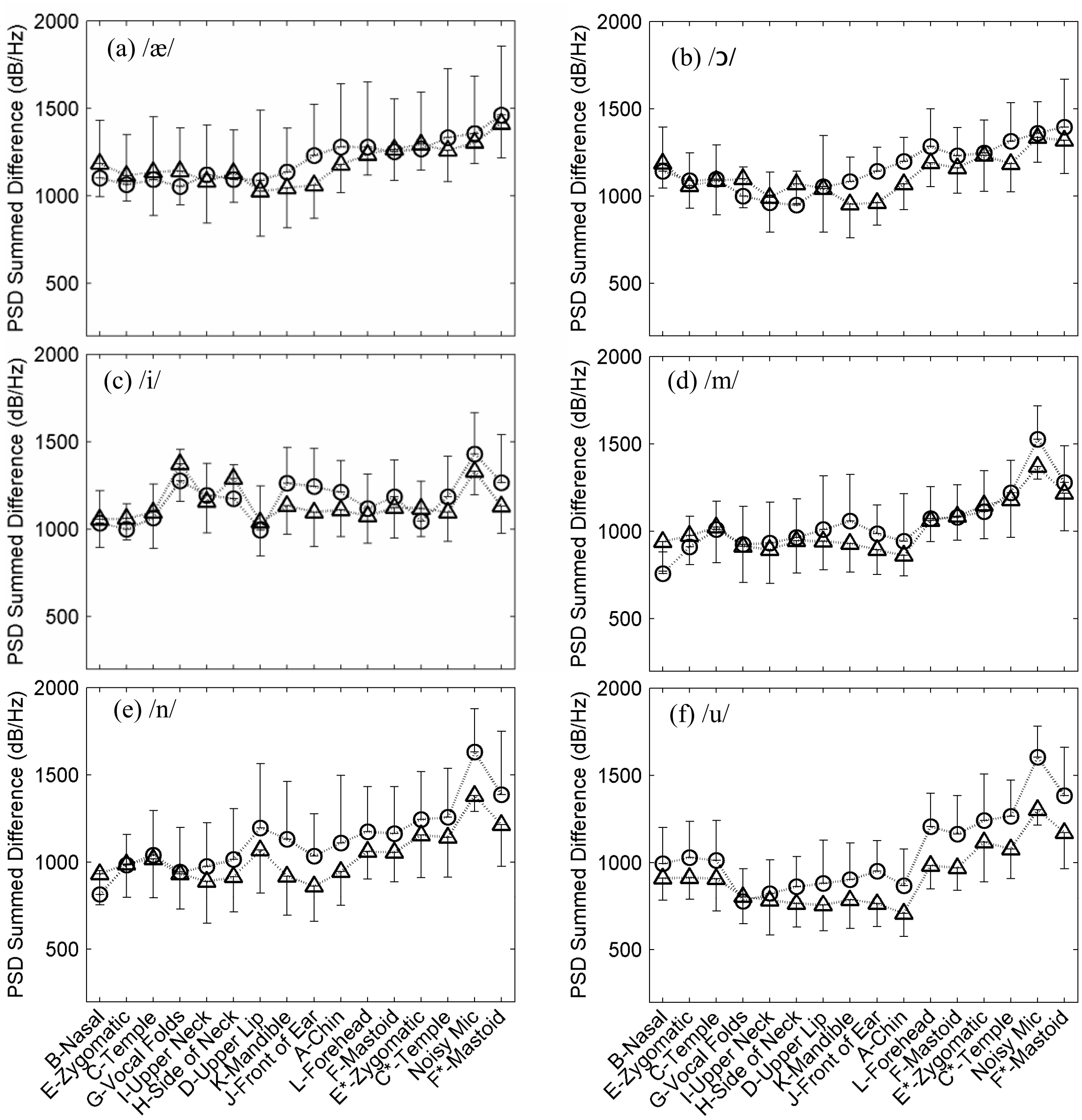

Figure 4-1 . Normalized PSD_{SD} over 0-5 kHz for phonation with background noise. ○: Male speakers; Δ: Female speakers. a) /æ/; b) /ɔ/; c) /i/; d) /m/; e) /n/; f) /u/.

4.2.1.2 Power Spectral Summed Difference Rankings with Noise

Tables 4.1 and 4.2 show the male and female average neck PSD_{SD} ranks, respectively. These tables are sorted according to the average rank over all sounds for each location. The tables also show the average rank for each sound at each location. Tables 4.1 and 4.2 also show that the only locations that are ranked in the top five for both male and female subject are the zygomatic (E) and the upper neck (I). These tables also show that the best overall ranked locations are different for male and female speakers. For both genders the noisy microphone ranked second worst.

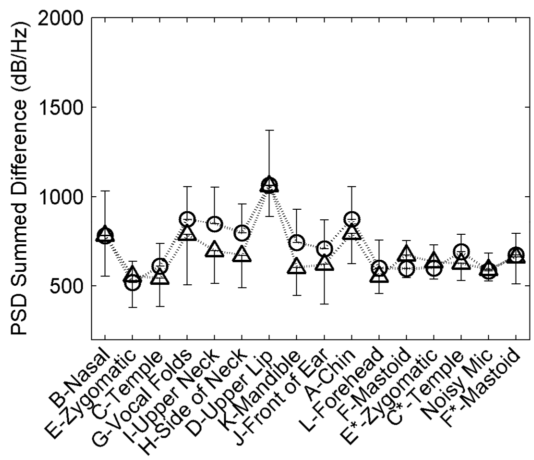

Figure 4-2 Sound /f/ normalized PSD_{SD} for sound production with background noise. ○: Male speakers; Δ: Female speakers.

4.2.2 Frequency Response with External Noise and No Phonation

4.2.2.1 Power Spectral Density of the Skin without Phonation for one Male Subject

Figure 4-3 shows the PSD analysis for one subject with external noise and no phonation. It is interesting to note that some locations had lower PSD values for the lower frequencies and higher PSD values for higher frequencies given roughly the same amplitude of input for each frequency. Figure 4-3 shows that for the male subject the chin, lip, over the vocal folds, the forehead, the zygomatic* process and the temple* had a nearly linear increase in PSD after 5 kHz.

Table 4-1 Average background PSD_{SD} subject rank, male speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
B	4.7	7.3	**3.2**	7.8	**2.2**	**1.9**	6.0	4.7
E	4.6	6.3	4.0	7.8	5.5	5.7	**2.7**	5.2
C	**4.5**	6.6	5.7	7.9	8.0	6.2	5.2	6.3
G	4.6	4.5	11.9	**2.7**	5.2	4.5	12.0	6.5
I	6.5	**3.7**	9.5	3.7	5.3	5.0	12.2	6.5
H	7.0	4.5	9.2	5.1	7.0	6.4	11.3	7.2
D	5.2	5.8	3.9	4.9	7.6	9.4	13.9	7.2
K	7.0	5.9	11.2	5.7	9.1	9.3	8.5	8.1
J	10.2	8.6	10.8	6.5	7.3	7.0	8.6	8.4
A	11.5	9.8	10.0	4.9	5.4	7.9	13.2	8.9
L	11.3	12.1	7.5	12.0	9.5	9.7	5.5	9.6
F	10.5	10.6	8.9	11.6	10.2	10.5	5.9	9.8
E*	11.5	11.5	5.0	12.0	10.8	11.2	6.7	9.8
C*	12.9	13.5	9.9	13.2	13.3	12.0	9.7	12.1
Mic	9.5	10.5	13.5	15.4	15.7	15.5	5.5	12.2
F*	14.5	14.6	11.8	14.6	14.1	13.9	9.0	13.2

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 4-2 Average background noise PSD_{SD} subject rank, female speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
J	4.56	**2.78**	7	4.11	4.44	**2.67**	7.22	4.7
K	**4.11**	**2.78**	9.78	4.89	5	4.33	6.22	5.3
I	5.56	4.44	9.11	4.89	4.33	3.44	10.1	6.0
A	9.11	6.44	7.44	**2.56**	**3.44**	5.67	13	6.8
E	6	6.11	6.33	9.11	7.33	8.22	4.89	6.9
D	**4.11**	6	**4**	3.44	5.89	10.3	15.8	7.1
H	7.78	7.11	13.7	4.78	6.44	4.89	8.89	7.7
C	8.22	7.44	7.33	8.89	10.1	9.67	**3.78**	7.9
B	8.33	10.8	**4**	10.6	7.67	7.22	8.89	8.2
G	7.11	8.67	14.9	6.89	5.22	6.11	11.2	8.6
L	10.1	11.3	6.22	10.4	10.4	9.67	4.78	9.0
F	11	10.7	9	10.4	11.1	9.89	8.89	10.1
C*	12.3	11.8	7.22	13.2	14.2	13.2	8	11.4
E*	12.7	12.8	9.11	13.7	13.3	13.3	7.89	11.8
Mic	9.56	11.8	11.9	14	14.1	14.3	7.33	11.9
F*	15.4	15.1	9	14.1	12.9	13	9.11	12.7

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Other locations, however, have relatively flat or even responses to the white noise input. The nasal bone, zygomatic, mastoid, angle of mandible, upper neck, and in front of the ear all have relatively flat PSD plots. This indicates that these locations are equally sensitive to external noise over the entire frequency range.

For this male subject Fig. 4-3 shows that for the zygomatic (E) and the side of the neck (H) the CC and OC configurations are similar at higher frequencies and the CO and OO configurations are also similar at higher frequencies. This indicates that for these locations, having the nasal passage open or closed noticeably influenced the response. However, the influence is different for each of the locations. For the zygomatic having the nasal passage open increased the PSD response at low fequencies, while for the side of the neck the open nasal passage resulted in a lower PSD response at low fequencies.

4.2.2.2 Average Power Spectral Density of the Skin without Phonation

Figure 4-4 shows the average PSD analysis over all subjects with external noise and no phonation. This figure shows that all locations have an increase in PSD with increasing frequency. The magnitude of this increase varies with location, ranging from around 10 dB to 30 dB.

Figure 4-4 shows that for the most part, having the mouth and nose open or closed did not make a large difference in the majority of the responses. However, there are some differences in the response with the varying configurations. For most locations there were differences in the PSD trends at lower frequencies. For the differences at lower frequencies the two mouth open configurations had similar responses and the two mouth closed configurations responses were similar, indicating that the sound transmission through the mouth has a greater effect on the response than through the nose. In each case when there is a difference at lower frequencies, the mouth open configurations resulted in higher PSD values. This indicates that having the mouth open transmits external low frequency noise through the vocal tract.

While for most locations the response at higher frequencies is similar for each configuration, the response at the forehead (L) and the nasal bone (B) have a distinct variation at higher frequencies. Figure 4-4 shows that for these locations at frequencies greater than 6 kHz the open nose configurations resulted in higher PSD values while the closed nose configurations had lower PSD values. This indicates that the forehead and the nasal bone are sensitive to high frequency noise introduced through the nasal passage.

Figure 4-4 shows that the nasal bone (B), mastoid (F), upper neck (I), and the forehead (L) have resonance spikes in the PSD plots at higher frequencies. The peaks all occur at different frequencies, although the forehead and nasal bone peaks occur with 1 kHz of each other. The peaks in the upper neck and the mastoid are preceded by anti-resonance dips while the nasal bone and the forehead only have resonance peaks.

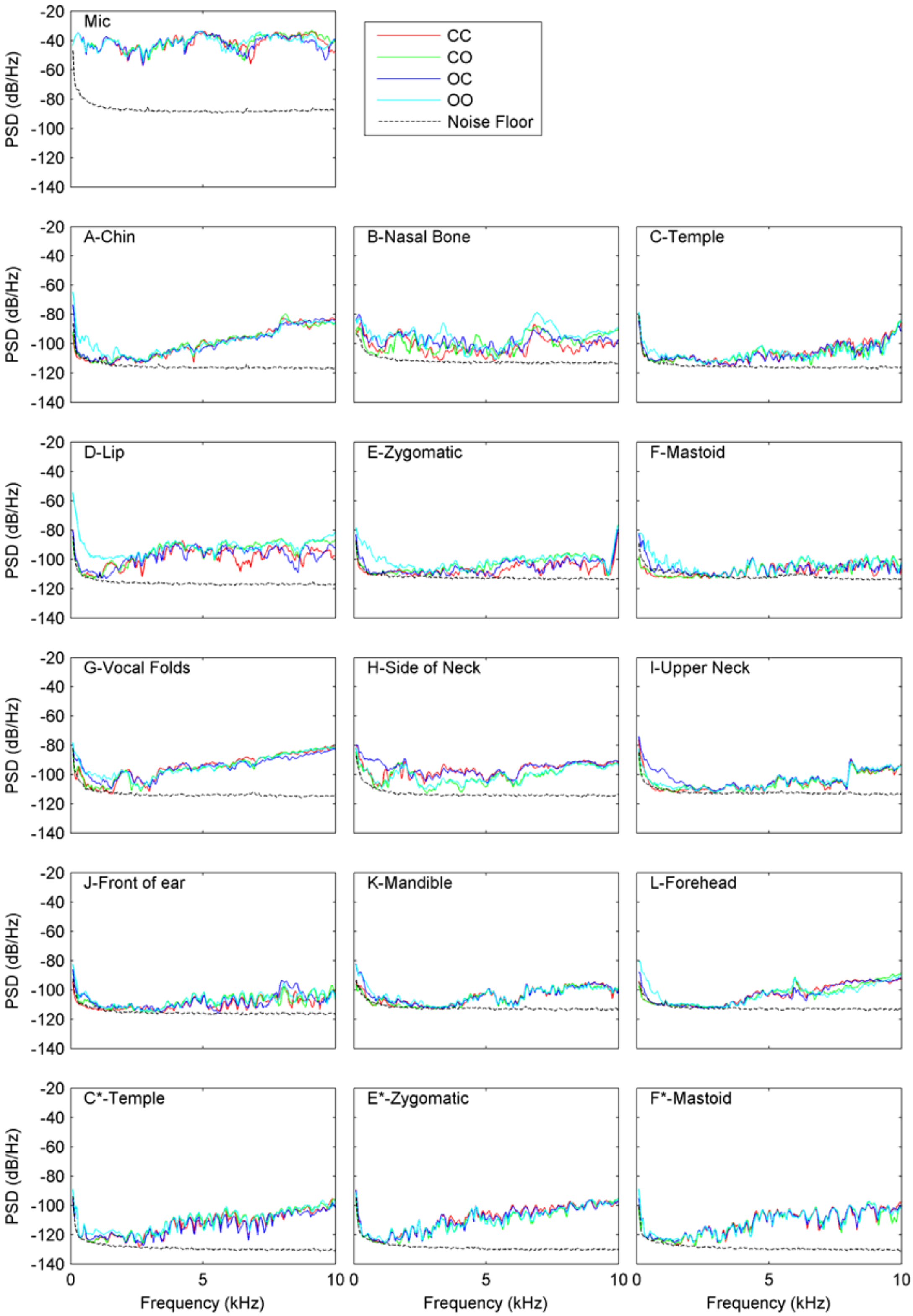

Figure 4-3 PSD plots for each location without phonation in the presence of 95 dB background noise for one male subject.

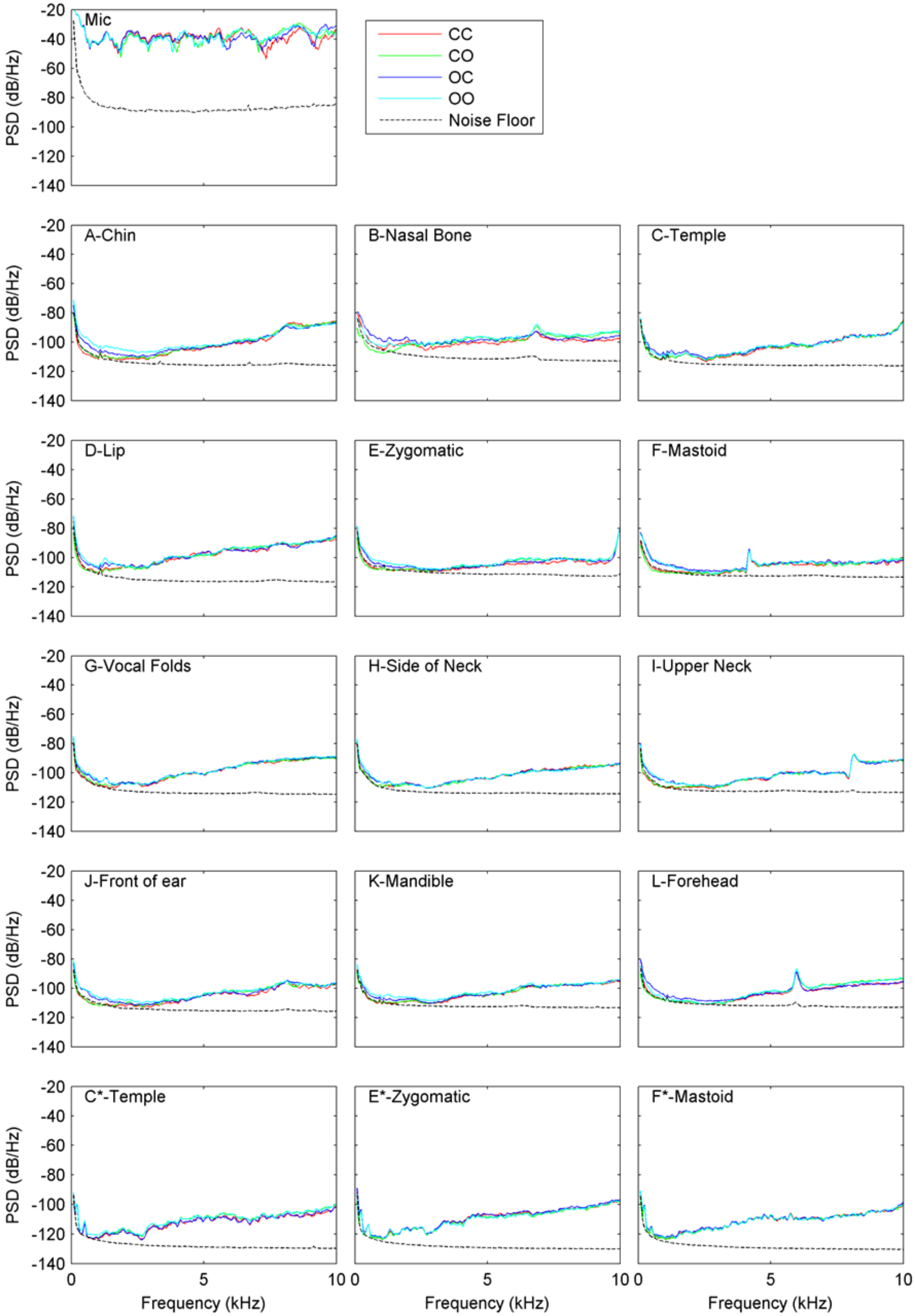

Figure 4-4 Average PSD plots for each location without phonation in the presence of 95 dB background noise.

4.3 Discussion of Results

4.3.1 Influence of Location

Figure 4.1 shows that for each sound there were many locations which had spectra that matched the clean microphone spectra better than the noisy microphone. This is more evident for sounds /i/, /u/, /m/ and /n/. For sounds /æ/ and /ɔ/ the noisy microphone had comparable PSD_{SD} values with a few locations. Tables 4.1 and 4.2 show that when averaged over all subjects and sounds, the noisy microphone ranked second to last. This indicates that in the presence of background noise, the accelerometerssensed a signal that better matched the clean microphone signal for most sounds and locations.

4.3.2 Difference in Gender

Table 4.1 shows that for male subjects, the locations that yield signals that best matched the clean microphone's spectra in the presence of background noise are the nasal bone, zygomatic, temple, over the vocal folds, and the upper neck. Table 4.2 shows that the best locations for female subjects are in front of the ear, the angle of mandible, the upper neck, the chin, and the zygomatic. Unlike in Chapter 3 (for no background noise), the top locations were not the same for both males and females. Only the zygomatic and the upper neck are ranked in the top five for both male and female subjects. It is somewhat surprising that the zygomatic ranked so well, given that the zygomatic has the second worst SNR for both male and female subjects (see Tables 5.3 and 5.4).

4.3.3 Clean vs. Noisy Power Spectral Density Summed Difference Ranks

In comparing the clean PSD_{SD} ranks in Table 3-4 with the noisy PSD_{SD} ranks in Table 4-1, we see that the nasal bone location (B) is the top ranked location in both tables. In comparing the tables we also see that the only locations that are in the top five for both clean and noisy environments are the nasal bone and the temple. This indicates that for male speakers, some locations may work well in both noisy and quiet environments.

However, when the clean and noisy PSD_{SD} ranks in Tables 3-5 and 4-2 are compared for female subjects, we find that there are no similarities in the top ranked locations. This may indicate that for females, accelerometer locations that work well in high noise environments may not work as well in quiet environments.

4.3.4 Frequency Response without Phonation

Figures 4-3 and 4-4 show that in the presence of external noise, low frequencies are attenuated and higher frequencies are transmitted at many locations. This trend is opposite of what happens to the frequency response of the skin in detecting speech. As seen in Fig. 3-2, in transmitting the speech signal, the skin and body tissues attenuate high frequencies from the vocal tract and pass low frequencies.

This information could be useful in designing acoustic shielding for contact microphones. It appears that the skin at some locations is less responsive to low frequency noise from external sources. Thus acoustic shielding at these locations may not need to provide as much low frequency attenuation. However, at some locations the skin is more sensitive to higher frequency external noise, indicating that the contact microphone at these locations should be designed to be shielded from high frequency noise.

4.4 Conclusions

This chapter includes an analysis of the frequency response of the skin on the face and neck in the presence of background noise with and without phonation. The conclusions are outlined as follows:

- During phonation in the presence of background noise, the accelerometers' signals better matched the clean microphone signal for most sounds and locations, supporting prior assertions that contact microphones may be better suited over traditional acoustic microphones for speech transmission in noisy environments.
- During phonation with background noise, the top ranked locations for male and female subjects had little agreement. Therefore different locations may be necessary to have adequate speech transmission for both genders.
- There is little agreement between the best ranked locations during phonation with and without background noise. Therefore, locations that work well in high noise environments may not work as well in quiet environments.
- Without phonation, the PSD plots show that low frequency signals from external noise are attenuated and higher frequency signals are transmitted. This indicates that contact microphones may be more sensitive to high frequency noise than low frequency noise.
- Having the mouth and nose open or closed did not make a large difference in the majority of the frequency responses.

- Without phonation, the skin frequency response is more sensitive to low frequency noise when the mouth is open.
- The forehead and the nasal bone are sensitive to high frequency noise introduced through the nasal passage.

5 Signal-to-Noise Ratio Analysis

An important consideration in selecting a transducer for transmitting speech is the amount of noise that accompanies the signal. Speech signals with high noise are typically of lower intelligibility and quality. Identifying transducers, locations, and filtering methods to reduce the transmitted noise is desirable to increase the clarity of speech in high noise environments. In this chapter we examine the signal-to-noise ratio (SNR) of speech signals detected on the skin surface in quiet and noisy environments. These SNRs are compared to the SNR of the microphone in the same environments.

The SNR data collection and analysis methods are described and the results are presented and discussed. In Chapter 3 the SNR results were presented for one sound; the Chapter 3 work is here expanded over all tested sounds and the results are generalized.

5.1 Methods

5.1.1 Experimental Setup

The same locations and procedures were used to collect the accelerometer and microphone signals as in Chapter 3. After completing the sounds and phrases in a quiet environment, as outlined in Section 3.1.1, the subjects were given ear plugs and repeated the same sounds and phrases in the presence of 95 dB white noise. A KRK Systems RP-6 studio monitor was placed 18 inches in front of the microphone and was used to play the white noise such that the measured sound pressure level at the microphone was 95 dB. The studio monitor had a reported frequency response of +/- 1.5 dB from 49 Hz to 20 kHz. Subjects 2, 6, 8, 9, and 23 had missing data for one or more of the accelerometers for the background noise portion of the testing so they were excluded from the SNR analysis.

5.1.2 Data Analysis

The signal-to-noise ratio (SNR) for the each sound was calculated for each subject and location with and without the presence of background noise. The SNR was calculated using the following equation:

$$SNR = 10\log_{10}\left(\frac{\frac{1}{k}\sum_{n=1}^{k}\left|signal(n)^2\right|}{\frac{1}{l}\sum_{m=1}^{l}\left|noise(m)^2\right|}\right), \tag{5.1}$$

where *signal* denotes a portion of the data during which the subject is talking, *noise* is a portion of the data prior to the subject talking, k is the number of data points in the signal portion, and l is the number of data points in the noise portion. The average SNR was also calculated for each location for all subjects.

5.2 Results

Figures 5-1 to 5-4 and Tables 5-1 to 5-4 show the average SNR for all sounds for male and female subjects with and without the presence of background noise. These data show that all locations are influenced somewhat by the presence of background noise, some considerably more than others. There are only minor differences between the male and female speakers, and the overall trends and values are very similar.

Tables 5-1 and 5-2 show that without background noise, the microphone has the best SNR, followed by locations on the nasal bone, chin, and throat. These tables show that most locations had SNRs above 10 dB without background noise. Tables 5-3 and 5-4 show that for both genders, the locations over the vocal folds, nasal bone, and upper neck had the best SNRs in the presence of background noise. While the microphone had the best SNR in a quiet environment (as would be expected), it had the third worst SNR out of all the locations in the presence of loud background noise.

Tables 5-5a-c are a summary of the SNR data for male and female subjects. Tables 5-5a and 5-5b are sorted according to their respective SNR with background noise and give the overall average SNR for each location with and without background noise. Table 5-5c shows the average change in SNR for each location. Table 5-5c shows that the microphone had the largest average decrease in SNR (about 32 dB), followed by the chin (about 12 to 13 dB), while locations in front of the ear, forehead, upper lip, nasal bone, and temple all had around a 9 to 10 dB decrease in SNR.

5.3 Discussion of Results

5.3.1 Sensitivity to Noise

Previous studies have emphasized the superior SNRs of contact microphones in high noise environments. We have found that this is the case for some locations, although other locations yielded generally poor SNRs. Figures 5-1 through 5-4 show that in quiet environments, all accelerometer locations had a SNR ratio much lower than the microphone signal, with the exception of the nasal sounds detected at the nasal bone, which were nearly 10 dB higher. However, with 95 dB background noise, Tables 5-3 and 5-4 show that, on average, all but 2 locations have higher SNR than the microphone, these locations being the two Zygomatic locations.

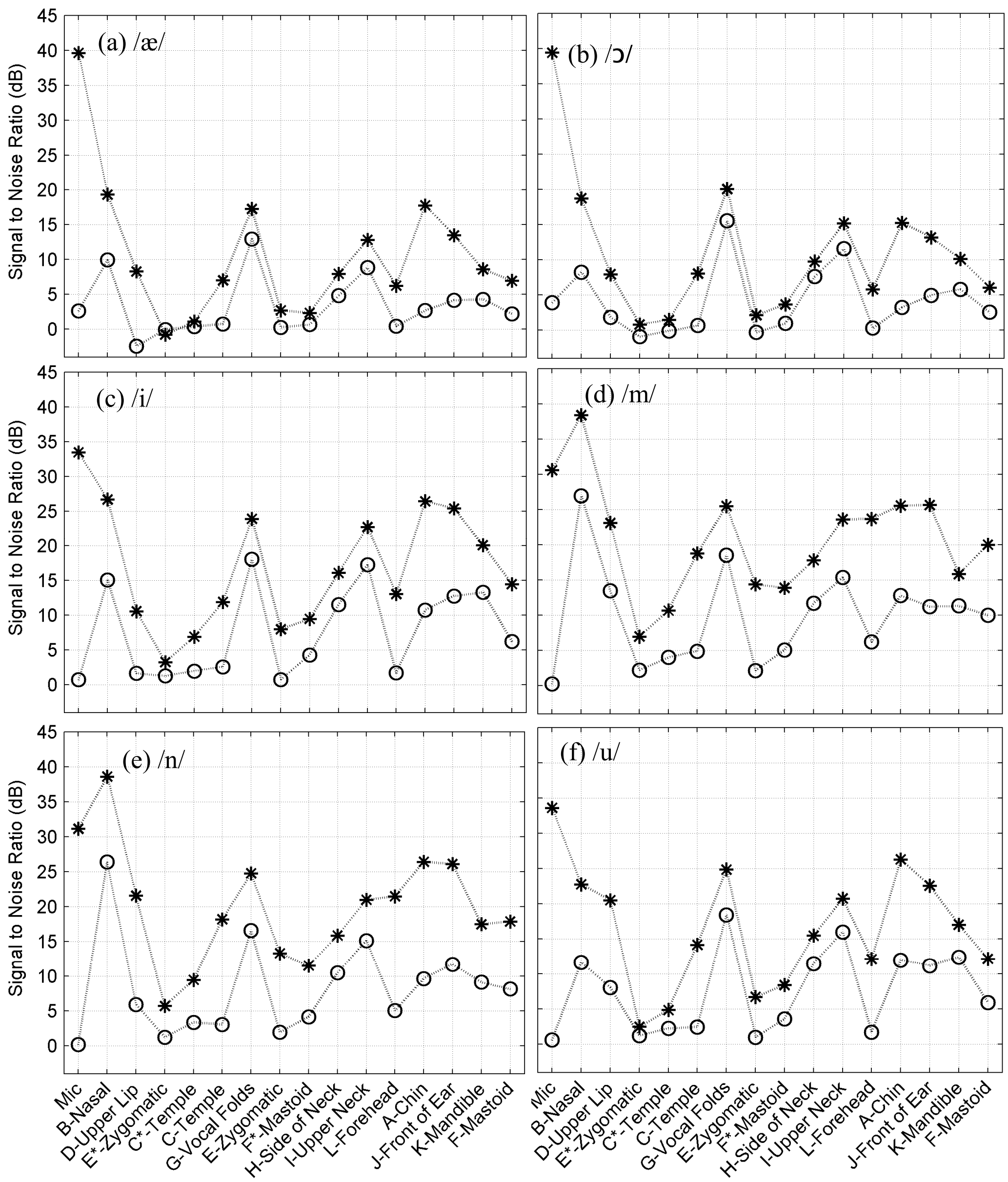

Figure 5-1. Average SNR for male subjects. ✱ No noise; ⊙ 95 dB background noise. a) /æ/; b) /ɔ/; c) /i/; d) /m/; e) /n/; f) /u/.

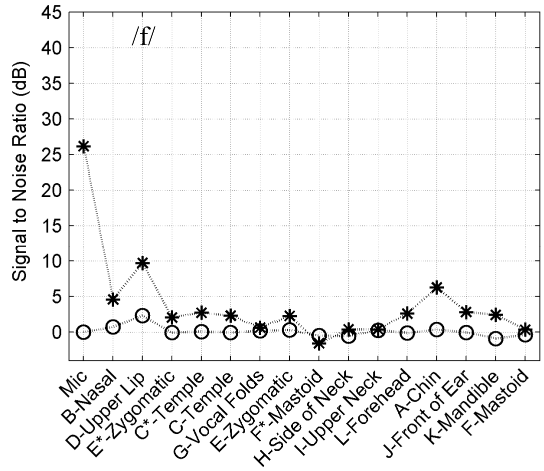

Figure 5-2 Sound /f/ average SNR for male subjects. * No noise; O 95 dB background noise.

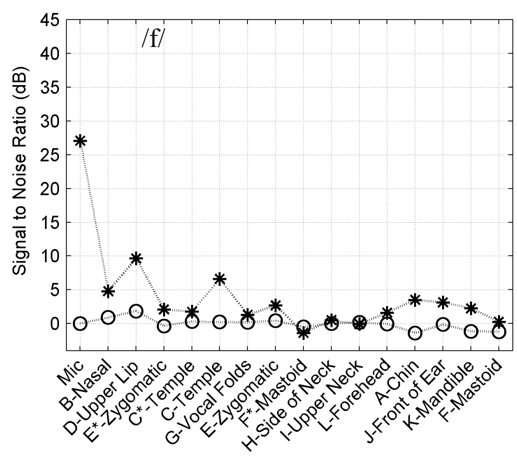

Figure 5-3 Sound /f/ average SNR for female subjects. * No noise; O 95 dB background noise.

While some locations were more sensitive to background noise than others, the microphone was more sensitive to the noise than any of the accelerometer locations. Table 5-5c shows that the microphone SNR for both male and female speakers decreased by about 32 dB with the 95 dB background noise. The most sensitive location to noise was the chin, which decreased about 13 dB for both male and female speakers. Five locations (front of the ear, forehead, upper lip, nasal bone and temple) experienced a reduction in SNR of about 9 to 10 dB.

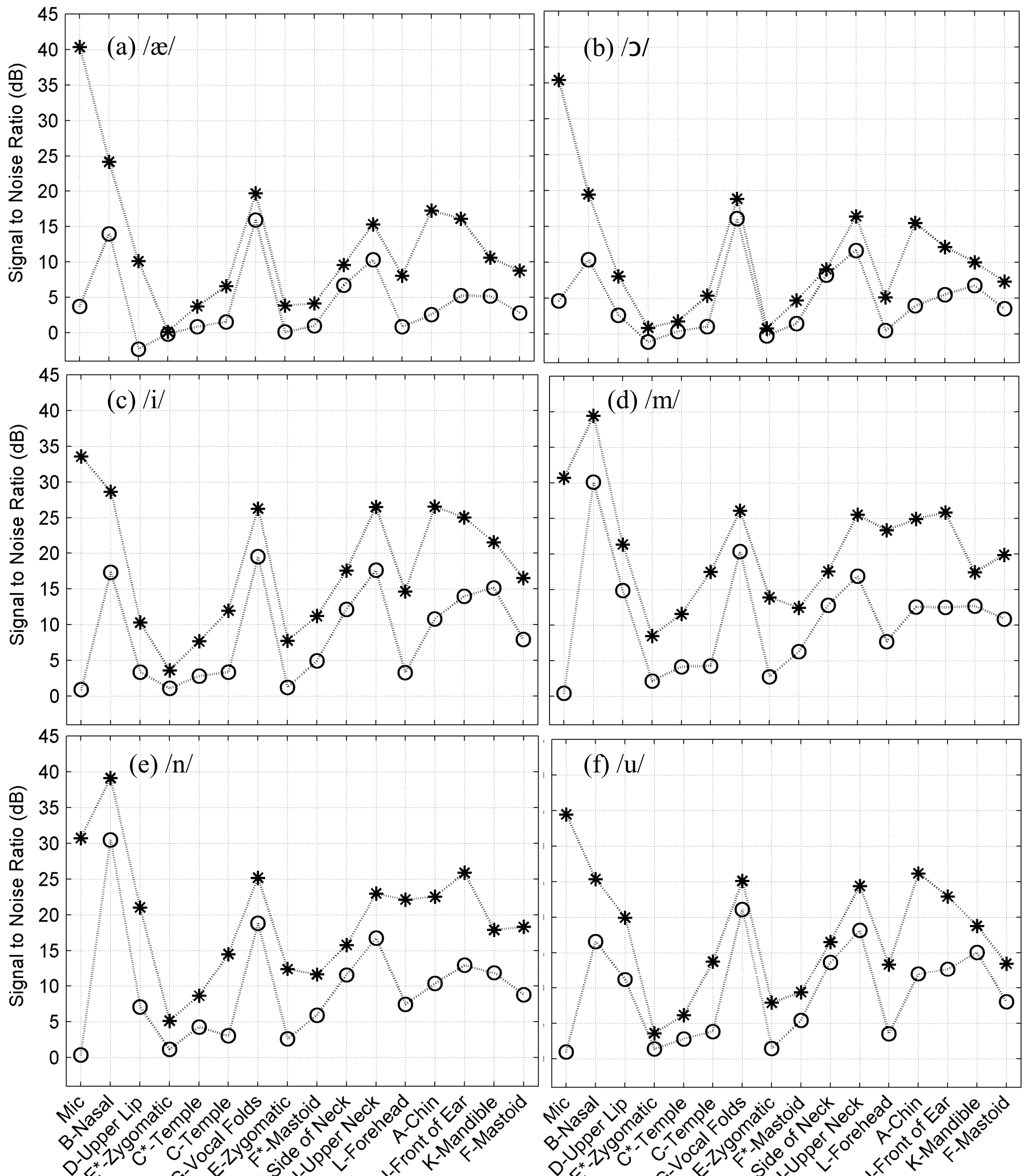

Figure 5-4 . Average SNR for female subjects. —*— No noise; ·○· 95 dB background noise. a) /æ/; b) /ɔ/; c) /i/; d) /m/; e) /n/; f) /u/.

Table 5-1 Average SNR without background noise, male speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
Mic	39.61	39.40	33.45	33.59	30.57	31.13	26.17	33.42
B	19.29	18.72	26.67	22.74	38.38	38.59	4.59	24.14
A	17.73	15.25	26.42	26.28	25.55	26.40	6.26	20.56
G	17.23	20.03	23.91	24.83	25.49	24.75	0.61	19.55
J	13.45	13.18	25.39	22.55	25.63	26.08	2.81	18.44
I	12.79	15.17	22.72	20.72	23.59	20.95	0.37	16.62
D	8.26	7.93	10.54	20.50	23.08	21.55	9.70	14.51
K	8.56	10.10	20.06	17.00	15.81	17.47	2.41	13.06
L	6.22	5.78	13.02	12.10	23.66	21.45	2.59	12.12
H	7.93	9.72	16.06	15.45	17.79	15.83	0.36	11.88
C	6.99	8.02	11.89	14.13	18.75	18.16	2.30	11.46
F	6.92	6.04	14.44	12.14	20.00	17.86	0.35	11.11
E	2.68	2.11	7.96	6.70	14.35	13.31	2.25	7.05
F*	2.33	3.64	9.45	8.39	13.88	11.51	-1.59	6.80
C*	1.06	1.43	6.89	4.85	10.65	9.45	2.72	5.29
E*	-0.79	0.79	3.25	2.47	6.95	5.71	2.05	2.92

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 5-2 Average SNR without background noise, female speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
Mic	40.32	35.38	33.55	34.44	30.72	30.74	27.06	33.17
B	24.17	19.46	28.66	25.34	39.40	39.11	4.77	25.85
G	19.65	18.84	26.27	25.09	26.07	25.17	1.29	20.34
A	17.24	15.46	26.54	26.15	24.94	22.52	3.49	19.48
J	16.09	12.10	25.04	22.90	25.84	25.89	3.12	18.71
I	15.31	16.40	26.49	24.39	25.50	22.97	-0.08	18.71
D	10.09	8.06	10.28	19.92	21.34	21.04	9.59	14.33
K	10.58	10.01	21.59	18.78	17.41	17.89	2.23	14.07
L	8.07	5.13	14.64	13.34	23.31	22.12	1.58	12.60
H	9.56	9.01	17.55	16.50	17.56	15.79	0.48	12.35
F	8.76	7.30	16.51	13.49	19.88	18.32	0.26	12.07
C	6.55	5.34	11.97	13.75	17.46	14.51	6.57	10.88
F*	4.12	4.66	11.18	9.38	12.44	11.62	-1.39	7.43
E	3.80	0.77	7.74	7.89	13.91	12.40	2.67	7.03
C*	3.74	1.75	7.66	6.16	11.56	8.64	1.75	5.89
E*	0.12	0.86	3.59	3.61	8.46	5.11	2.08	3.40

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 5-3 Average SNR with 95 dB background noise, male speakers.

Location	Sounds /æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	Average
G	12.92	15.51	18.02	18.42	18.50	16.55	0.14	14.30
B	9.92	8.24	15.02	11.65	26.95	26.35	0.71	14.12
I	8.82	11.55	17.22	15.98	15.34	15.13	0.27	12.04
H	4.78	7.58	11.53	11.48	11.67	10.47	-0.57	8.13
J	4.13	4.95	12.75	11.16	11.19	11.72	-0.05	7.98
K	4.29	5.79	13.30	12.37	11.31	9.12	-0.91	7.90
A	2.65	3.25	10.72	11.95	12.81	9.65	0.33	7.34
F	2.17	2.55	6.18	5.92	10.00	8.15	-0.37	4.94
D	-2.45	1.84	1.63	8.06	13.43	5.91	2.31	4.39
F*	0.65	0.97	4.28	3.56	5.03	4.16	-0.52	2.59
L	0.40	0.28	1.69	1.70	6.20	5.05	-0.13	2.17
C	0.74	0.64	2.56	2.44	4.88	3.06	-0.06	2.04
C*	0.38	-0.14	1.95	2.23	4.00	3.37	0.07	1.69
Mic	2.61	3.91	0.71	0.60	0.26	0.20	0.00	1.18
E	0.24	-0.34	0.74	0.99	2.14	1.92	0.28	0.85
E*	-0.06	-0.92	1.24	1.19	2.20	1.22	-0.08	0.68

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 5-4 Average SNR with 95 dB background noise, female speakers.

Location	Sounds /æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	Average
B	13.95	10.34	17.33	16.58	30.07	30.49	0.88	17.09
G	15.89	16.08	19.47	21.08	20.33	18.80	0.18	15.97
I	10.31	11.66	17.58	18.13	16.86	16.74	0.15	13.06
K	5.19	6.73	15.14	15.05	12.67	11.86	-1.15	9.35
H	6.67	8.21	12.14	13.62	12.79	11.56	-0.01	9.28
J	5.24	5.48	13.92	12.68	12.50	12.97	-0.15	8.95
A	2.53	3.95	10.79	12.03	12.56	10.33	-1.40	7.25
F	2.81	3.52	7.93	8.02	10.86	8.77	-1.22	5.81
D	-2.32	2.64	3.35	11.20	14.88	7.07	1.83	5.52
F*	0.97	1.47	4.91	5.39	6.24	5.93	-0.50	3.49
L	0.82	0.47	3.30	3.53	7.69	7.43	-0.08	3.31
C	1.53	1.01	3.34	3.85	4.26	3.01	0.21	2.46
C*	0.84	0.36	2.80	2.80	4.15	4.28	0.28	2.22
Mic	3.70	4.61	0.88	0.99	0.41	0.33	0.01	1.56
E	0.10	-0.23	1.20	1.45	2.75	2.60	0.40	1.18
E*	-0.22	-1.13	1.07	1.39	2.13	1.14	-0.37	0.57

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastoid process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastoid process	I - Upper neck	L - Forehead

Table 5-5 SNR summary for male and female speakers with and without added background noise. a) Male SNR comparison with and without background noise, b) Female SNR comparison with and without background noise, c) Average drop in SNR from no noise cases to 95 dB noise cases, all values dB.

a)

Location	Male SNR Clean	Male SNR Noisy
G	19.6	14.3
B	24.1	14.1
I	16.6	12.0
H	11.9	8.1
J	18.4	8.0
K	13.1	7.9
A	20.6	7.3
F	11.1	4.9
D	14.5	4.4
F*	6.8	2.6
L	12.1	2.2
C	11.5	2.0
C*	5.3	1.7
Mic	33.4	1.2
E	7.1	0.9
E*	2.9	0.7

b)

Location	Female SNR Clean	Female SNR Noisy
B	25.8	17.1
G	20.3	16.0
I	18.7	13.1
K	14.1	9.4
H	12.4	9.3
J	18.7	9.0
A	19.5	7.3
F	12.1	5.8
D	14.3	5.5
F*	7.4	3.5
L	12.6	3.3
C	10.9	2.5
C*	5.9	2.2
Mic	33.2	1.6
E	7.0	1.2
E*	3.4	0.6

c)

Location	SNR Change Male	SNR Change Female
Mic	-32.2	-31.6
A	-13.2	-12.2
J	-10.5	-9.8
L	-9.9	-9.3
D	-10.1	-8.8
B	-10.0	-8.8
C	-9.4	-8.4
F	-6.2	-6.3
E	-6.2	-5.8
I	-4.6	-5.6
K	-5.2	-4.7
G	-5.3	-4.4
F*	-4.2	-3.9
C*	-3.6	-3.7
H	-3.7	-3.1
E*	-2.2	-2.8

A - Chin, B - Nasal Bone, C - Temple, C* - Temple, D - Upper Lip, E - Zygomatic, E* - Zygomatic, F - Mastoid process, F* - Mastoid process, G - Over vocal folds, H - Side of neck, I - Upper neck, J - Front of Ear, K - Angle of mandible, L - Forehead

The locations least affected by additional background noise were all of the higher sensitivity accelerometers, located at the mastoid, temple and zygomatic, and the side of the neck. While the higher sensitivity accelerometers did not experience a large reduction in SNR with the addition of noise, their no noise SNRs were relatively low. This is attributed to the generally small vibration amplitudes at these locations.

The side of the neck, on the other hand, had a 10 dB SNR without noise and only decreased between 3 and 4 dB when noise was added for both male and female speakers. The other neck locations, located over the vocal folds and the upper neck, also performed well, only decreasing between 4 and 6 dB with the added noise. These two locations may have had a higher reduction in SNR due to the fact that they were on the front of the neck directly incident to the white noise source, while the accelerometer placed on the side of the neck was not. While the locations on the upper neck and over the vocal folds had a larger reduction in SNR, they still resulted in a higher SNR in the presence of background noise than did the accelerometer on the side of the neck. This indicates that while a given location may be less sensitive to noise or have a smaller reduction in the SNR in the presence of noise, other locations that may be more sensitive to noise or have larger reductions in SNR may still yield signals with higher SNRs.

As these tests were done with white background noise, they represent a generalized case, and many high noise environments will not necessarily have flat spectra. If a particular operating environment is know to have background noise with a specific spectrum, it is recommended that studies be conducted to identify locations and filtering methods that result in clear speech for those particular environments.

Another consideration is that the noise level was constant during testing. Operating conditions may have noise levels that are louder or softer than the test conditions presented. The effect of varying noise level was not explored. The noise level may influence different locations differently due to the various composition and non-linear properties of the underlying tissue. It is thus recommended that the effects of varying noise level be explored.

5.3.2 SNR Performance

Tables 5.1 and 5.2 show that for both male and female subjects, the areas with the highest SNR in the quiet environment are the nasal bone, over the vocal folds, the chin, the angle of mandible, and the upper neck. In the presence of background noise, the locations with the highest SNR for both male and female subjects are over the vocal folds, the nasal bone, upper neck, the side of the neck, in front of the ear, and the angle of mandible. The locations that had the highest SNRs in both quite and noisy environments were over the vocal folds, nasal bone, upper neck, and in front of the ear. It is intuitive that the neck locations would have favorable SNRs owing to their close proximity to the vibration source. Additionally, the location in front of the ear is located on the posterior portion of the zygomatic arch, which is bony process with little tissue covering. The nasal bone location is also a bony structure with little tissue covering. The good SNR performance in both quiet and noisy environments for the nasal bone and the location in front of the ear could be due to bone conduction of the speech signal.

5.3.3 Response of Different Phonemes

In a quiet environment, Tables 5-1 and 5-2 show that the nasal bone has the best SNR for all vowels and nasals, with the following three 3 exceptions. For male speakers the SNR for the sound /ɔ/ was highest over the vocal folds. For both the male and female subjects the SNR for the sound /u/ was highest at the chin. For both the male and female speakers the sound /ɔ/ had a comparable SNR at the nasal bone and over the vocal folds. Both of these areas have good SNRs for all sounds, with the nasal bone typically having higher SNRs. The chin also had good SNRs for all sounds in a quiet environment.

As reported in Chapter 3, the nasal bone also was one of the locations that matched the microphone spectra best in quiet environments and was rated well by listeners. Although the locations over the vocal folds, chin, and in front of the ear have good SNRs, they did not have good PSD_{SD} values and did not rate well with listeners.

This indicates that even though an area has a good SNR it does not mean that the signal is of good quality. Signals that were rated high by listeners had both high SNRs and adequate high frequency content, while signals that were rated as being lower quality had low SNRs and were more muffled. While the upper neck and upper lip did not have as good an SNR as these other locations, they were still between 14 and 18 dB. The upper neck had an average PSD_{SD} rank of 7.4 for female speakers, but 9.6 for male speakers. The upper lip, however, had good PSD_{SD} values for both male and female speakers. Both the upper lip and upper neck locations were rated well by listeners.

In the presence of background noise, the vowels tested had the highest SNRs at the location over the vocal folds and the nasal sounds had the highest SNRs on the nasal bone. This is consistent with the data presented in Chapter 3 (see Figure 3-6) that indicate that nasal sounds are amplified on the nasal bone and that the neck has high skin vibration amplitudes. Only the locations over the vocal folds, the nasal bone, and upper neck had SNRs above 10 dB in the noisy environment. The side of the neck, in front of the ear, the angle of mandible, and the chin had SNRs between 7 and 9 dB. The other nine other locations had SNRs lower than 6 dB. In comparison there were only two locations (*zygomatic and *temple) that had a SNR less than 6 dB in the quiet environment, and two others (*mastoid and zygomatic) that had SNRs under 10 dB for both male and female subjects. This shows that while many of the locations may have adequate SNRs in quiet environments, the reduction in SNR in high noise environments may result in signals that are inadequate for clear speech transmission. It is important to note that the SNR is only one consideration in selecting an ideal location; transmitted spectra, the overall intelligibility, and listener preference are also important considerations.

For all instances /f/ had the highest SNR on the upper lip. The sound /f/ is formed by turbulent airflow passing the teeth and lips in the front of the mouth. The accelerometer placed on the upper lip is directly over the sound source. However, it is interesting to note that for both male and female speakers, the upper lip had the worst PSD_{SD} rank of all the locations tested.

In quiet environments, the nasal bone and chin, comparatively speaking, had high SNRs for the sound /f/. The nasal bone appears to be a good location for bone conduction of this fricative. Table 3-6 shows that the nasal bone had the best PSD_{SD} rank for female subjects and Table 3-5 shows that the nasal bone had the 7th best PSD_{SD} rank for male subjects. The nasal bone also adequately matches the frequency spectra for the fricative /f/.

5.3.4 Considerations

An important consideration to note is that the accelerometers used in this study were not directional or shielded from external acoustical energy. The drop in SNR seen in many of the locations may be minimized by implementing a contact microphone with its external surface acoustically shielded from the surrounding environment. Further, matching the impedance between the transducer and the skin surface may improve the noise-eliminating ability of contact microphones. Application of pressure on the accelerometer on the skin may also improve the SNR.

5.4 Conclusions

In this chapter the SNR of the skin vibration signal was investigated with and without the presence of background noise. The conclusions reached are outlined as follows:

- All locations are influenced somewhat by external noise, some considerably more than others. This suggests that using contact microphones at some locations may result in favorable SNRs, but at other locations contact microphones may not have adequate SNRs.
- Without background noise the locations with the best SNR for both male and female subjects were the nasal bone, over the vocal folds, the chin, the angle of mandible, and the upper neck.
- With background noise the locations with the highest SNR for both male and female subjects were over the vocal folds, the nasal bone, the upper neck, the side of the neck, in front of the ear, and the angle of mandible.
- Locations that yielded the best SNR in both noisy and quiet environments were over the vocal folds, the nasal bone, in front of the ear, and the angle of mandible.
- The microphone had the best SNR with no background noise, but had the third worst SNR in the presence of 95 dB white noise.
- The microphone was more sensitive to noise than any of the accelerometers and had a 32 dB reduction in SNR when noise was added.
- Some locations yielded a good SNR but poor PSD_{SD} and perceptual ratings, thus indicating that a good SNR does not directly correlate with the quality of signal.

6 Neck Frequency Response Characterization

In some situations it may not be possible to wear or place a contact microphone on the face. If the neck is the only option for microphone placement, it is desirable for it to be located where the frequency response is the best. In this chapter the data collection and analysis methods to obtain the frequency response of the skin around a concentrated area on the neck during speech production are described. Results are reported for power spectral density summed difference (PSD_{SD}) values and rankings.

6.1 Methods

6.1.1 Experimental Setup

To test the frequency response of the skin on the neck during speech, accelerometers were attached to 12 locations on the neck of three male and four female subjects using medical-grade double-sided adhesive tape (see Fig. 6-1). The males had an average age of 24.7 years and the females had an average age of 24 years. One subject reported having speech therapy in elementary school; all other subjects reported having no history of voice or speech problems. All testing was done with IRB approval and in accordance with IRB policies. Prior to accelerometer placement the subjects removed oil and/or makeup with an alcohol prep pad to ensure adequate adhesion.

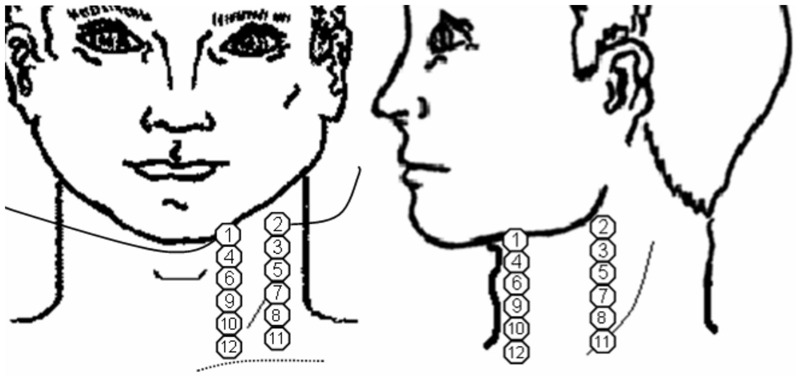

Figure 6-1 Accelerometer placement locations for neck location testing (image courtesy U.S. Army Research Lab Human Research & Engineering Directorate).

All accelerometers were manufactured by PCB Piezotronics; model numbers and relevant specifications are listed in Table 6-1. All accelerometers were placed on the left side of the neck. The subjects held a tape measure next to their neck and digital images were acquired of the accelerometer placement locations (see Fig. A-2). The wires for all accelerometers were attached to a head rest to minimize the torque on the skin due to the weight of the wires. The pressure of the accelerometers on the skin was not measured. All other experimental procedures were the same as outlined in Section 3.1.1.

6.1.2 Data Analysis

The same data analysis procedures as outlined in Section 3.1.2 were used to calculate the PSD_{SD} and PSD_{SD} ranks for the neck location signals. In addition to the PSD_{SD} values, the *x* and *y* location of each accelerometer was calculated from the digital images. In Figs. 6-4 to 6-11 the origin (0,0) corresponds to the superior anterior notch of the thyroid cartilage ("Adam's Apple"). The digital images of the accelerometer placement locations were imported in to MATLAB, and the 'ginput' command was used to collect the *x* and *y* pixel location of the thyroid cartilage, all 12 accelerometers, and a reference distance on a tape measure held next to the subjects neck while the image was taken (see Fig. A-2). The *x* and *y* pixel locations were then converted into centimeters using the reference distance to give the approximate distances relative to the thyroid cartilage for each subject. Since only one image was used to extract the position, the *x* and *y* positions are not exact, but instead represent approximate locations.

Table 6-1 Accelerometer locations and specifications (locations identified in Fig. 6-1).

Locations	Accelerometer Model #	Mass [g]	Sensitivity [mV/(m/s^2)]	Frequency range [Hz] (+- 5%)
1-4	352A24	0.8	10.2	1 to 8000
5-12	352A56	1.8	10.2	0.5 to 10000

6.2 Results and Discussion

6.2.1 Power Spectral Density Summed Difference

Figures 6-2(a-f) show the average PSD_{SD} over 0-5 kHz for each of the neck locations for male and female speakers for vowel and nasal sounds. These figures show that, generally, the PSD_{SD} increases toward the lower neck. Recall that a reduction in PSD_{SD} indicates a signal spectrum that better matches the microphone spectrum. The trends in the figures are similar; however, for males there is generally a “dip” in the PSD_{SD} values from locations 5 to 6, while for females this generally occurs from locations 3 to 4. There are a few outliers of interest. For sounds /u/, /m/ and /n/, location 9 had a reduction in PSD_{SD} that is not present in the other sounds. For the sound /i/, on average, female speakers showed a large reduction in PSD_{SD} for locations 8 and 11, while males showed a reduction for location 11.

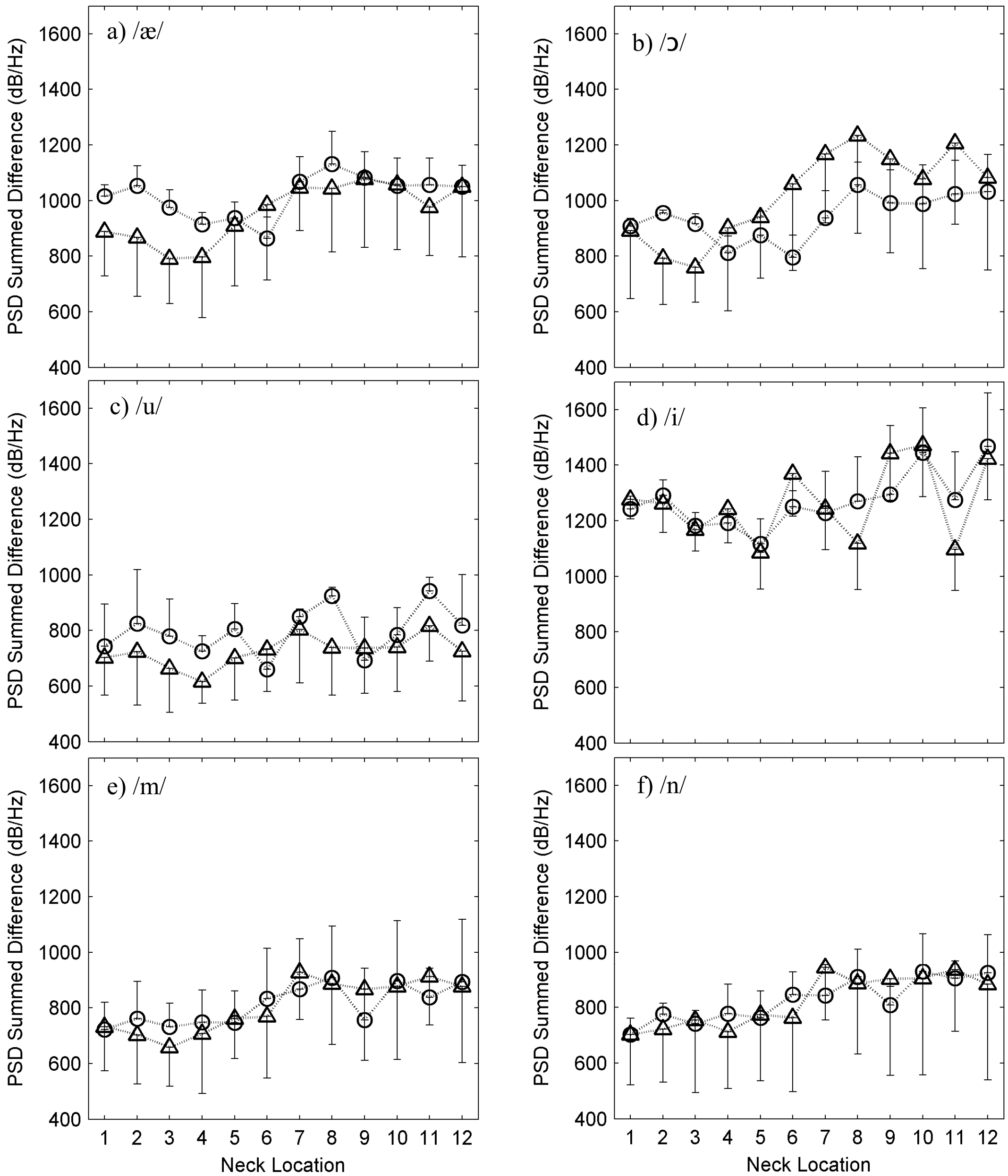

Figure 6-2 Normalized PSD_{SD} over 0-5 kHz for neck locations. ○: Male speakers; Δ: Female speakers. a) /æ/; b) /ɔ/; c) /u/; d) /i/; e) /m/; f) /n/.

Figure 6-3 shows the PSD_{SD} for the sound /f/. This figure indicates that for male subjects, on average, the locations that best match the microphone spectra are 8, 7, and 5. For female speakers the locations that best match the microphone spectra are 8, 5, and 11. For both male and female speakers these locations have average PSD_{SD} values much lower than the other 9 locations (see Tables 6-2 and 6-3).

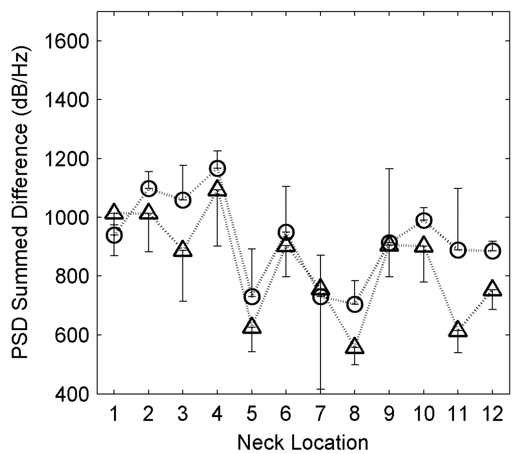

Figure 6-3 . Sound /f/ normalized PSD_{SD} for neck locations. ○: Male speakers; Δ: Female speakers.

Figure 6-2 plots the PSD_{SD} vs. location and gives an indication of how each location performed compared to the other locations. For most sounds, the locations that performed the best (had the lowest PSD_{SD} values), had average PSD_{SD} values 200-300 dB/Hz lower than the locations that had the highest PSD_{SD} values. This difference is comparable, but lower than the difference in the PSD_{SD} plots shown in Figs. 3-3(a-f), which have differences between 300-600 dB/Hz for the locations on the face and neck. This indicates that there is not as great a difference between the PSD_{SD} of the neck locations as there is between the neck and face locations.

Figure 6-2(c) shows the vowel sound /u/ has low PSD_{SD} values for most of the neck locations, indicating that it matches the microphone well on the neck. It is also seen in Fig. 6-2(a) that the vowel sound /i/ has PSD_{SD} values much greater than the other sounds, indicating that it is not detected very well on the neck. The trends seen in Fig. 6-2 indicate that locations lower on the neck generally have higher average PSD_{SD} values, but this figure also shows that the standard deviation is fairly high for many of the sounds and locations. This variation is attributed to the small sample size, and it is recommended that future studies include a larger number of subjects to verify these results and better locating of positions.

Figures 6-4 to 6-11 show the PSD_{SD} vs. position for each accelerometer for each subject. The color represents the PSD_{SD} level, with darker levels indicating better matching of the accelerometer and microphone spectra. These figures illustrate the differences in PSD_{SD} for each subject, as well as for the different sounds. As shown in Figure 6-6, the vowel /i/ for all subjects generally had higher PSD_{SD} values than all of the other sounds, indicating that the skin vibration signal spectra for /i/ matched the microphone spectrum worse than for the other sounds. Figure 6-7 shows the opposite for the sound /u/.

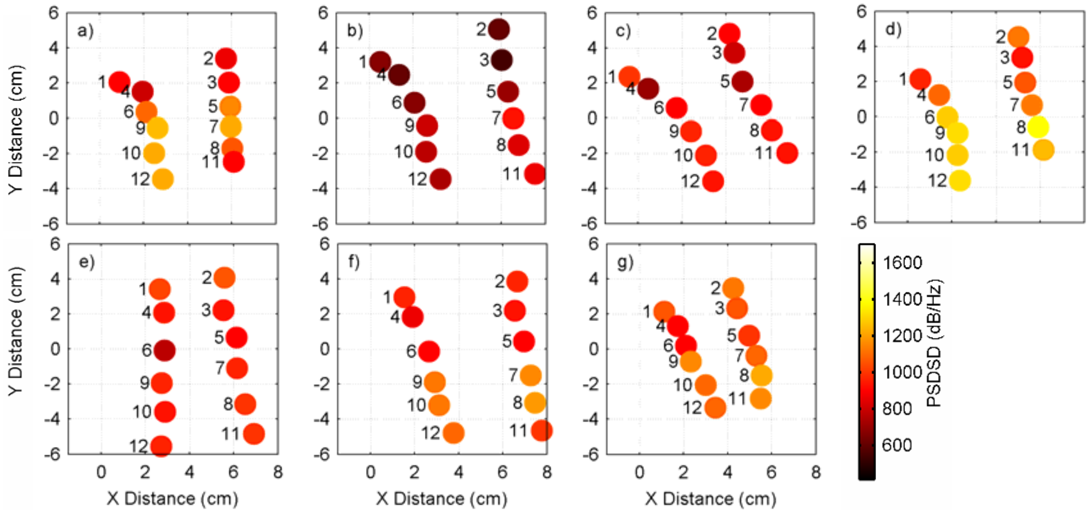

Figure 6-4 . Sound /æ/ neck location PSD_{SD} plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

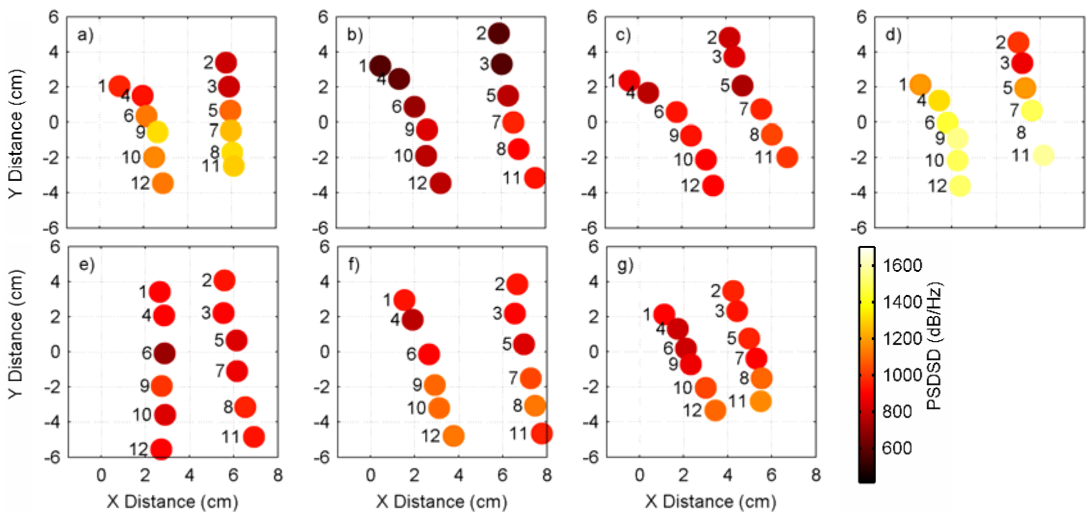

Figure 6-5 . Sound /ɔ / neck location PSD_{SD} plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

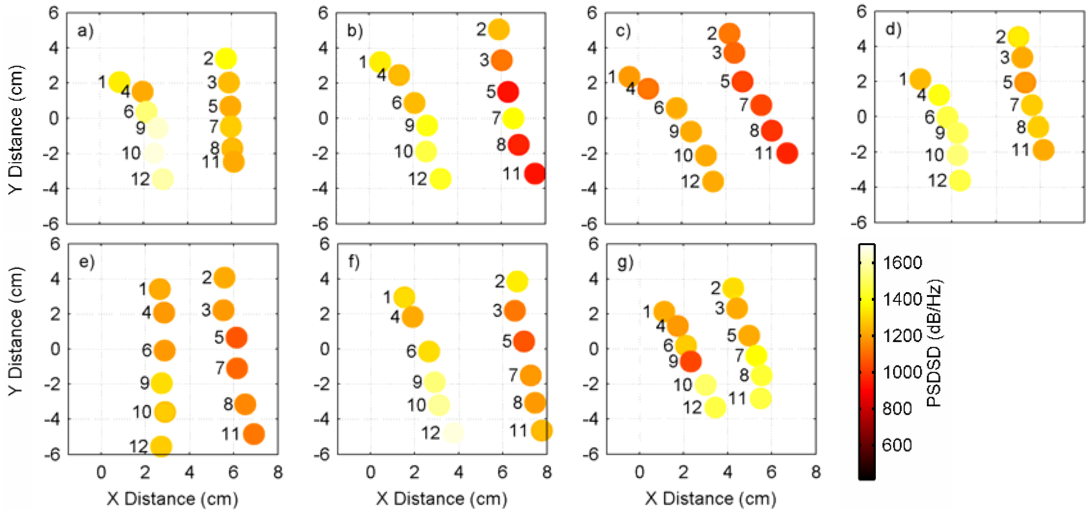

Figure 6-6 Sound /i/ neck location PSD$_{SD}$ plots. Female subjects a) 26; b) 27; c) 29; d) 32. Male subjects e) 25; f) 30; g) 31.

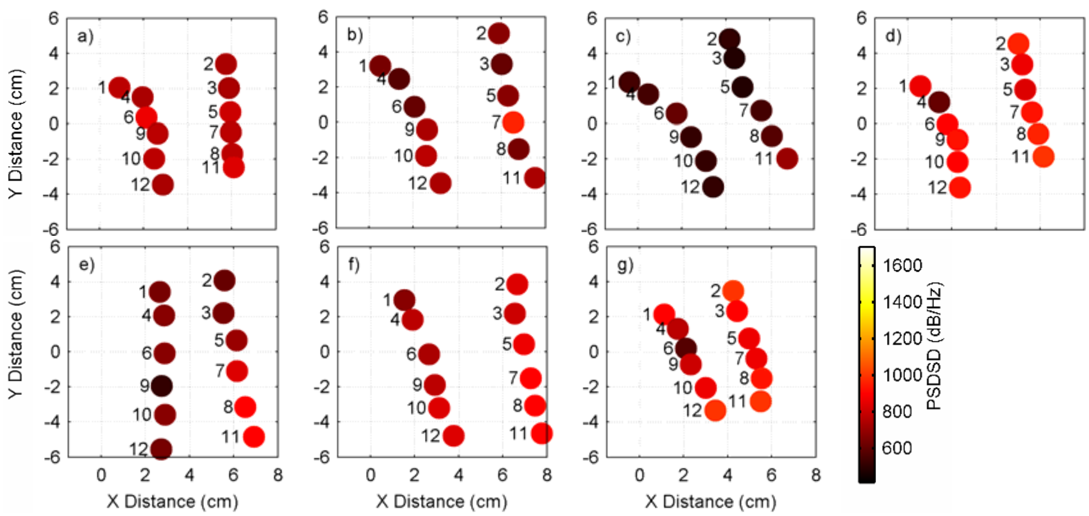

Figure 6-7 Sound /u/ neck location PSD$_{SD}$ plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

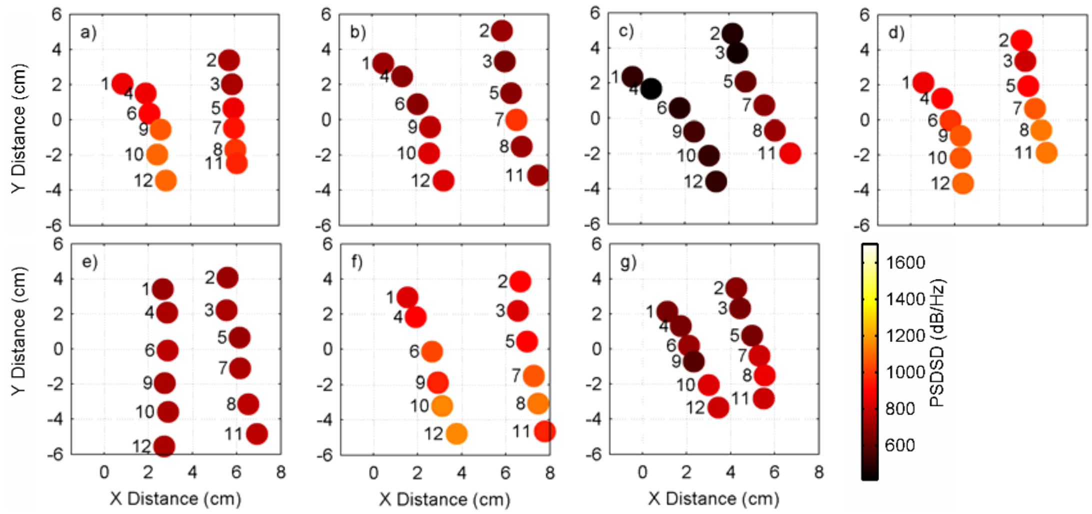

Figure 6-8 Sound /m/ neck location PSD_{SD} plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

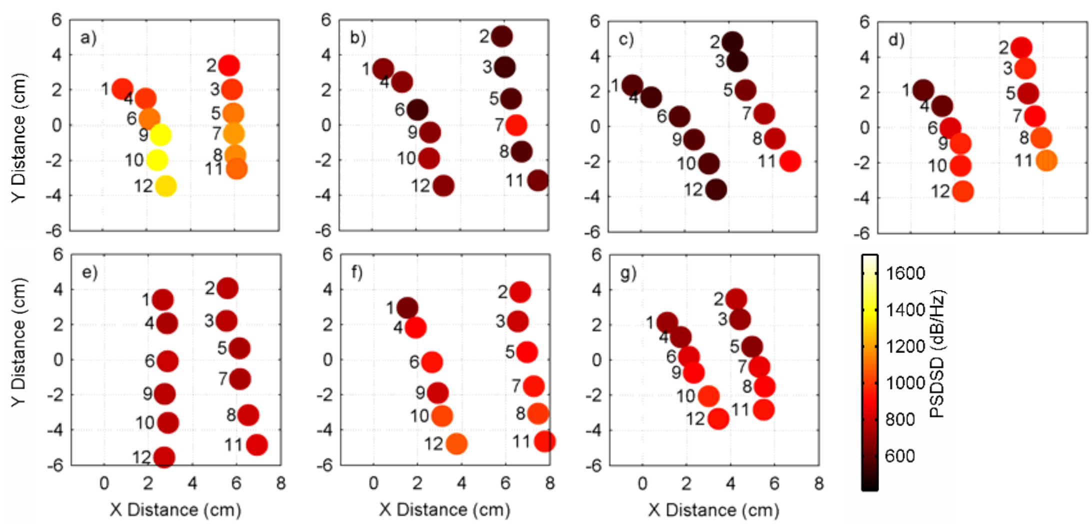

Figure 6-9 Sound /n/ neck location PSD_{SD} plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

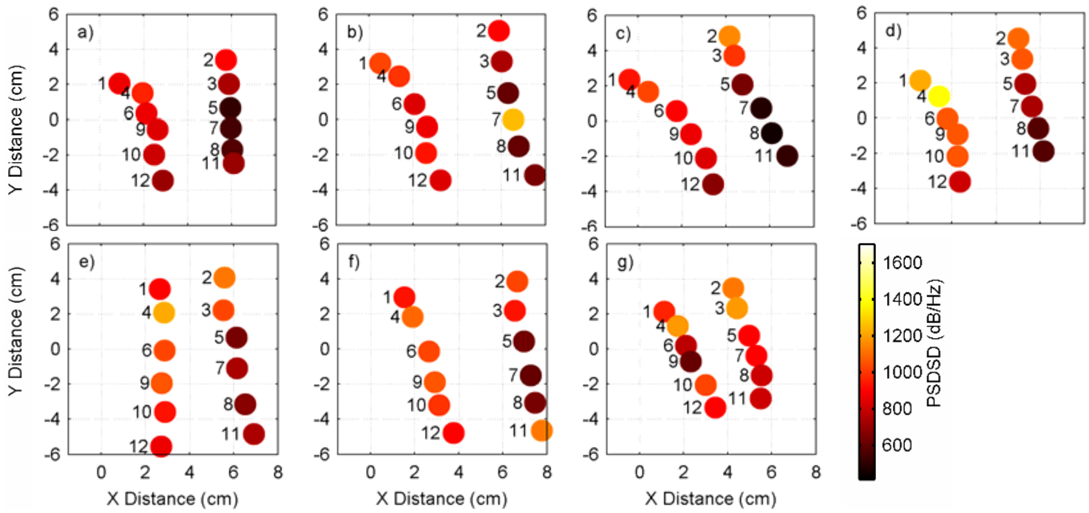

Figure 6-10 Sound /f/ neck location PSD_{SD} plots. Female subjects a) 26; b) 27; c) 29; d) 32; Male subjects e) 25; f) 30; g) 31.

6.2.2 Power Spectral Density Ranking

Tables 6-2 and 6-3 show the male and female average PSD_{SD} ranks, respectively. These tables are sorted according to the average rank over all sounds for each location. The tables also show the average rank for each sound at each location. These tables show that for the male subjects, locations 5 and 4 yielded the best average ranking. Locations 6, 3, and 1 all had similar average rankings, with location 3 having relatively flat or consistent rankings while locations 6 and 1 have a range of rankings over the various sounds. For female subjects, the best locations were 3, 5, and 4. Locations on the upper neck generally ranked better than those on the lower neck.

Table 6-2 Average neck PSD_{SD} rank, male speakers.

Neck Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
5	2.33	4.33	**1.67**	8.00	4.00	**2.67**	3.33	3.76
4	2.67	2.67	4.00	3.67	3.67	4.00	11.33	4.57
6	**1.67**	**2.00**	6.67	3.33	8.33	7.33	6.33	5.10
3	4.00	5.67	4.67	5.33	3.67	3.33	9.00	5.10
1	7.00	6.00	6.33	4.67	**1.67**	4.00	6.33	5.14
9	9.00	8.00	7.67	**2.67**	3.67	5.00	7.00	6.14
7	8.00	5.33	4.33	8.00	9.00	5.67	3.00	6.19
2	9.00	7.67	8.33	7.33	4.33	4.67	9.67	7.29
10	7.00	7.00	11.00	6.33	9.67	10.33	7.67	8.43
12	7.00	9.00	11.33	6.67	10.00	10.67	5.00	8.52
11	9.00	9.33	6.33	11.33	9.00	10.33	6.33	8.81
8	11.33	11.00	5.67	10.67	11.00	10.00	**3.00**	8.95

Table 6-3 Average neck PSD_{SD} rank, female speakers.

Neck Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
3	**2.00**	**1.75**	4.00	**2.00**	**1.25**	**3.50**	7.25	3.11
5	4.75	4.50	**1.75**	5.25	5.25	5.25	2.50	4.18
4	2.25	3.50	5.00	3.50	2.75	4.25	11.25	4.64
2	3.25	2.50	7.25	6.00	3.50	3.25	10.25	5.14
1	6.00	3.50	6.75	6.00	4.50	4.25	10.25	5.89
6	6.25	6.50	9.25	7.75	5.00	4.75	7.50	6.71
8	9.00	11.25	3.75	8.00	9.50	8.50	**1.75**	7.39
11	7.25	10.75	2.25	11.25	9.75	8.75	2.75	7.54
12	8.75	7.25	10.00	5.75	9.00	8.25	5.00	7.71
7	8.50	9.50	6.00	8.50	9.50	9.25	5.00	8.04
10	9.25	7.25	11.25	7.25	9.25	8.75	7.00	8.57
9	10.75	9.75	10.75	6.75	8.75	9.25	7.50	9.07

Tables 6-2 and 6-3 also show that locations on the lower neck generally have higher or worse PSD_{SD} rankings than the upper neck locations. However, locations 1 and 2 which are located at the top of the neck (just under the jaw) have worse PSD_{SD} ranks than the locations immediately below them (3, 4, 5). This can also be seen by examining

Figs. 6-4 to 6-11. This indicates that accelerometers placed above the thyroid cartilage typically have spectra that match the microphone spectrum better than accelerometers placed below the superior notch of the thyroid cartilage. This also indicates placing accelerometers too high on the neck may also lead to signals that do not match the microphone's spectrum as well as accelerometers placed a little lower on the neck.

The locations higher on the neck are further away from the vocal folds and have more tissue between them and the vocal tract. This distance away from the vocal tract and the increased amount of tissue likely contributes to the decrease in ranking for these locations. The locations in the middle of the neck are still near the sound source, but are also a little closer to the oral cavity than the locations on the lower neck. The higher PSD_{SD} values of the locations on the upper middle of the neck are attributed to their proximity to both the sound source and the oral cavity, where the higher frequency vowel sounds and consonants are shaped.

6.2.3 Difference Between Male and Female Speakers

For both male and female speakers, the mid-upper neck resulted in signals that better matched the spectra of the microphone. Location 5 was the top ranked location for male subjects while location 3 was the top ranked location for the female subjects. Both locations 3 and 5 are located on the side of the neck, with 5 being just under 3. For the male subjects the second best ranked location is location 4 which is on the front upper portion of the neck. For female subjects the second best ranked location was location 5 which is just below location 3 on the side of the neck.

An interesting result is that location 6, located just laterally to the thyroid notch where many current throat microphones are placed, had the third highest ranking for males and the sixth highest ranking for females. This supports the conclusions of Section 3.3.2 that indicate that a throat microphone placed close to the vocal folds may work better for male speakers than for female speakers. Since location 6 was not a top-ranked location, this indicates that there are locations that may be better suited for contact microphone placement than over the thyroid cartilage, even if the neck is the preferred location for microphone placement.

6.2.4 Differences in Sounds

Location 3 ranked best, on average, for female speakers for all sounds except /i/ and /f/. The sound /i/ was ranked best at location 5 while /f/ was ranked best at location 8 on the lower side of the neck. However, when listening to the recorded data from location 8, the fricative sound /f/ was inaudible. Thus this result is attributed to the accelerometer noise that has a spectrum that matches the relatively flat response of the microphone, as discussed in Section 3.3.3.

For both male and female subjects, the sound /i/ was ranked best at location 5. This is the only sound that had the same best ranked location for both the male and female subjects. It is also interesting to note that location 3, which ranked the best for 5 sounds for the female subjects, was not the top ranked location for any of the sounds for the male subjects.

For the male subjects the top locations were much less consistent for the various sounds. The sounds /æ/ and /ɔ/ were best ranked at location 6. The sounds /i/ and /n/ were best ranked at location 5. The sound /m/ was best ranked at location 1. The sound /u/ was best ranked at location 9. Due to the high variation in the top ranked locations for the male subjects in this study it is recommended that further investigation with a larger study population be made to determine the best overall neck location for male speakers.

For both male and female speakers, the top three locations for the sound /f/ were the same. The accelerometer at these locations, however, did not seem to be sensing the speech sound, but rather seemed to just be transmitting noise. For the fricative sound /f/, a perceptual rating should be used to determine the best microphone location.

6.3 Conclusions

In this chapter the results from studying the frequency response of the neck were presented and discussed. The conclusions reached are outlined below:

- There are locations that may be better suited for contact microphone placement other than directly over the thyroid cartilage (where many throat microphones are currently placed) if the neck is the preferred location for microphone placement.
- Generally the PSD_{SD} increases towards the lower locations, corresponding to poor matching of accelerometer and microphone spectra.
- For both male and female speakers the upper middle portion of the neck had the best PSD_{SD} rankings. For the male subjects locations 5, 4 and 6 yielded the best average ranking. For female subjects the best ranked locations were 3, 5 and 4.

It is important to note that filtering may be needed to reestablish attenuated high frequency content and to obtain adequate intelligibility if the signal is only detected at the neck.

7 Conclusions and Recommendations

The objective of this thesis work has been to assist in the development of improved contact microphones and associated signal processing methods by contributing to an improved understanding of the frequency response of the skin at various locations on the head and neck during speech. This chapter summarizes the conclusions reached as a result of this research and gives recommendations for future work.

7.1 Head and Neck Frequency Response Characterization

In Chapter 3, the speech signal detected on the skin was characterized by attaching accelerometers to the skin on the head and neck and having subjects produce various sounds and a phrase. The accelerometer and microphone PSDs were compared in order to identify locations that yield an accurate representation of the audible speech signal. A perceptual listening test was conducted to compare the PSD rankings to perceptual rankings. The following are the main conclusions reached in this portion of the study:

- Locations other than on the throat can adequatelysense the speech signal from the skin. Some of these locations were found to yield signals with spectra better matching that of the microphone and that yielded higher perceptual ratings than the throat signal.

- While the throat attenuates the signal intensity of high frequencies, some high frequency content can be detected at other locations on the head.
- Using PSD_{SD} as the metric, the locations that best matched the frequency content of the microphone are generally common to both male and females. These locations are the nasal bone, above the upper lip, both temple locations, and the zygomatic*.
- Using perceptual ratings as a metric, the highest rated locations are the nasal bone and the zygomatic* followed by both temple locations, above the upper lip, and the upper neck.
- Perceptual rankings generally follow the PSD_{SD} ranking, with a few outliers. This indicates that locations' signals that match the spectrum of the microphone are also generally (but not always) preferred by listeners.

7.2 Frequency Response Characterization with External Noise

In Chapter 4 the analysis done in Chapter 3 was repeated with the addition of background noise. The noisy accelerometer signals were compared to the clean microphone signals in order to identify locations that were least sensitive to noise by producing signals similar to the clean microphone signal even in the presence of noise. With background noise, the frequency response of the skin was also found without phonation. The following are the main conclusions reached as a result of this analysis.

- During phonation in the presence of background noise, the accelerometers detected a signal that better matched the clean microphone signal for most sounds and locations, supporting prior assertions that contact microphones may be better suited over traditional acoustic microphones for speech transmission in noisy environments.

- During phonation with background noise, the top ranked locations for male and female subjects did not correspond with one another. Different locations may therefore be necessary to enable adequate speech transmission for both genders.
- There is little agreement between the best ranked locations during phonation with and without background noise. Thus locations that work well in high noise environments may not work as well in quiet environments.
- Without phonation, the PSD plots show that low frequency vibration signals from external noise are attenuated and higher frequency signals are transmitted. This indicates that contact microphones may be more sensitive to high frequency noise than low frequency noise.

7.3 Signal-to-Noise Analysis

In Chapter 5, the SNR of the skin vibration signal was investigated with and without the presence of background noise. The main conclusions reached are outlined as follows:

- All locations are influenced somewhat by external noise, some considerably more than others. This suggests that using contact microphones at some locations may result in favorable SNRs, but at other locations contact microphones may not have adequate SNRs.
- Locations that yielded the best SNR in both noisy and quiet environments were over the vocal folds, the nasal bone, in front of the ear, and the angle of mandible.

- Some locations yielded favorable SNRs but yielded poor PSD_{SD} and perceptual ratings, thus indicating that a good SNR does not directly correlate with a high quality signal.

7.4 Neck Frequency Response

In Chapter 6 the frequency response of the skin around a concentrated area on the neck was analyzed. This was done in order to gain a better understanding of the signals on the neck, if the neck is the only available region for speech detection. The following are the main conclusions from this analysis:

- There are locations that may be better suited for contact microphone placement other than directly over the thyroid cartilage (where many throat microphones are currently placed) if the neck is the preferred location for microphone placement.
- For both male and female speakers the upper middle portion of the neck had the best PSD_{SD} rankings.

Note that given the small sample size (7 subjects) and the amount of variation in some of the data, it is recommended that these observed trends be verified with a study involving more subjects.

7.5 Optimal Location

During the analysis one location did not perform the "best" in all circumstances. However, a few locations performed well in many of the analyses; the nasal bone was one of these locations. It ranked well in both the PSD_{SD} analysis as well as the perceptual rankings in quiet environments. It had the best overall PSD_{SD} rank for male subjects with background noise and overall had favorable SNRs. This location, however, may not be the most convenient location to place a contact microphone. There are many factors to consider in identifying the "optimal" location for a contact microphone. The "optimal" location for microphone placement should:

- Produce a signal that adequately matches the spectrum of a clean microphone signal in both quiet and noisy environments,
- Produce a signal that has an adequate SNR in both quiet and noisy environments, and
- Be in a location for convenient placement.

7.6 Recommendations

Since the coupling between the skin and the accelerometers was not controlled, the preferred contact microphone locations identified in this study should be treated with some caution. It is recommended that future studies quantify the effects of contact pressure on the skin frequency response at the various locations. Future work is also suggested to analyze more phonemes (including more fricatives and stop sounds) and phrases.

Some locations have adequate SNRs and are in locations that are convenient for mounting, but do not adequately match the spectra of the clean microphone signal. It is recommended that future work investigate the potential for formulating transfer functions and signal processing methods that will restore the lost/attenuated frequency content to improve speech quality for all sounds at various locations.

It is also recommended that, although very time intensive, perceptual studies be conducted whenever possible to verify the analytical (e.g., PSD_{SD}) conclusions reached. A quality method to use in these perceptual studies is the Modified Rhyme Test which compares similar sounding words for intelligibility. This will ensure that the locations and signals identified as preferable are actually producing signals that are preferred by the human ear. It is recommended that perceptual studies also determine what PSD_{SD} values correspond with acceptable intelligibility and/or quality.

During the study the same type of transducer was used at each location, and no attempt was made to improve sound transmission via impedance matching. The impedance of areas above soft tissue, such as on the neck, may be very different from the impedance of tissue over bony structures. This is a limitation of the current study and it is recommended that further research investigate the implementation of devices that match the impedance of the various areas.

Appendix A

Table A-1 Phonetic symbols of sounds tested.

Symbol	Sound
/æ/	h***a***t
/ɔ/	***ough***t
/i/	f***ee***t
/m/	***m***an
/n/	***n***ut
/u/	b***oo***t
/f/	***f***ast

Table A-2 Average background noise PSD_{SD}, male speakers.

Location	Sounds							Average
	/æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	
B	1062	1130	997	1047	**726**	**794**	630	912
E	1063	1087	998	1029	910	983	**518**	941
G	**1052**	1000	1276	**777**	924	943	874	978
I	1120	962	1194	821	932	976	847	979
H	1094	**948**	1174	863	964	1016	796	979
C	1092	1096	1064	1012	1009	1040	609	989
D	1088	1052	994	882	1010	1196	1063	1041
J	1231	1142	1243	952	987	1035	706	1042
K	1136	1083	1263	902	1058	1130	742	1045
A	1279	1197	1211	868	942	1108	873	1068
F	1251	1231	1185	1164	1079	1163	597	1096
L	1278	1283	1118	1205	1072	1174	600	1104
E*	1269	1247	1044	1240	1112	1245	600	1108
C*	1334	1314	1184	1266	1220	1256	689	1180
Mic	1231	1233	1318	1447	1395	1497	588	1244
F*	1461	1395	1266	1384	1278	1387	672	1263

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastiod process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastiod process	I - Upper neck	L - Forehead

Table A-3 Average background noise PSD_{SD}, female speakers.

Location	Sounds /æ/	/ɔ/	/i/	/u/	/m/	/n/	/f/	Average
J	1062	961	1096	762	894	**861**	621	894
K	1045	**952**	1133	786	929	919	601	909
I	1082	992	1159	783	896	888	696	928
A	1180	1069	1109	**708**	**861**	944	792	952
E	1115	1060	1061	913	975	990	556	953
H	1130	1070	1290	764	946	916	669	969
C	1138	1087	1097	907	1024	1019	**542**	973
D	**1026**	1043	1040	754	944	1068	1062	991
B	1153	1186	**1035**	982	971	975	675	997
G	1141	1098	1374	804	913	932	786	1007
L	1237	1190	1074	984	1061	1060	554	1023
F	1266	1161	1123	969	1084	1055	672	1047
C*	1260	1184	1096	1078	1180	1142	625	1081
E*	1294	1233	1119	1118	1149	1155	633	1100
Mic	1204	1199	1212	1132	1226	1236	588	1114
F*	1413	1319	1131	1171	1217	1214	662	1161

A - Chin	C* - Temple	E* - Zygomatic	G - Over vocal folds	J - Front of Ear
B - Nasal Bone	D - Upper Lip	F - Mastiod process	H - Side of neck	K - Angle of mandible
C - Temple	E - Zygomatic	F* - Mastiod process	I - Upper neck	L - Forehead

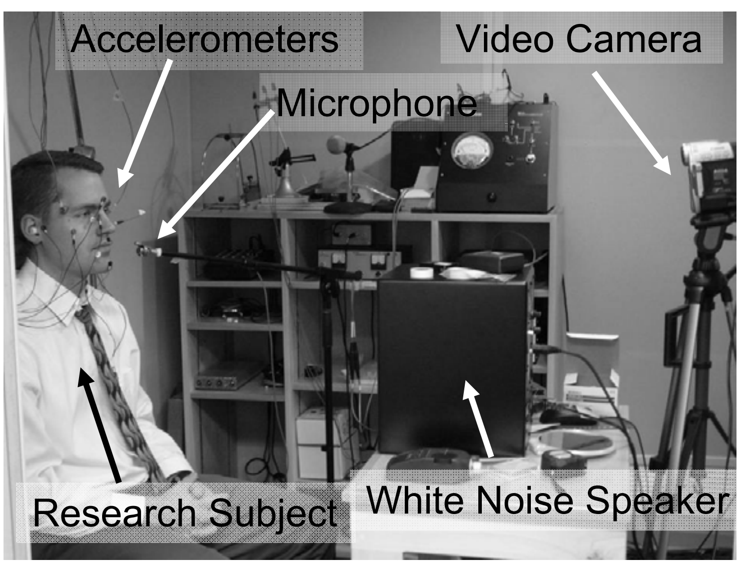

Figure A-1 Sound booth testing setup.

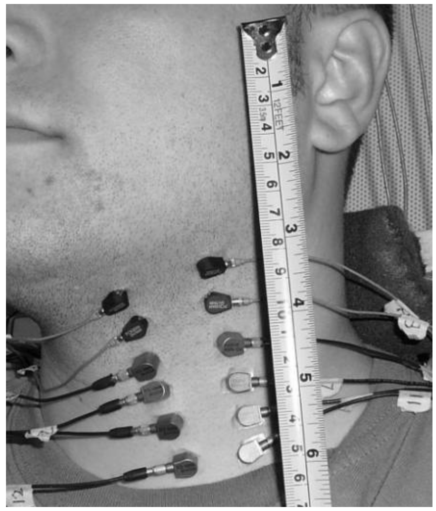

Figure A-2 Example of the neck placement location digital images for chapter 6.

References

Acker-Mills, B. E., Houtsma, A. J. M., and Ahroon. W. A. (2004). "Speech intelligibility in noise using throat and acoustic microphones," USAARL Report No. 2004-13.

Cheyne, H. A. II, (2002). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Ph. D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Cheyne, H. A., Hanson, H. M., Genereux. R. P., Stevens K. N., Hillman, R. E. (2003). "Development and testing of a portable vocal accumulator" Journal of Speech, Language, and Hearing Research **46**:1457-1467, December 2003.

Dromey, C., Nissen, S., Roy, N., Merril, R. (2008). "Articulatory changes following treatment of muscle tension dysphonia: Preliminary acoustic evidence" Journal of Speech, Language, and Hearing Research **51**:196-208, February 2008.

Dupont, S., Christophe, R., Bachelart, D. (2004). "Combined use of close-talk and throat microphones for improved speech recognition under non-stationary background noise," Proceedings of Robust 2004 (Workshop (ITRW) on Robustness Issues in Conversational Interaction), Norwich, Aug. 2004.

Garciarena, M., Franco, H., Sonmez, K., Bratt, H,. (2003). "Combining standard and throat microphones for robust speech recognition," IEEE Signal Processing Letters **10**(3):72-74.

Hayes, D. P., Meltzer, L. (1967). "Bone-conducting microphones," The American Journal of Psychology **80** (4):619-624.

Horacek, J., Vesely, J., Pesek, L., Vohradnik, M. (2004). "Fundamental dynamic characteristics of human skull: Part I – Experimental modal analysis and FE modeling of basic vibration properties," Engineering MECHANICS **11**(2):139-158.

IEEE, (1969). IEEE recommended practice for speech quality measurements (IEEE No. 297). New York: IEEE.

Meltzner, G. S., Kobler, J. B., Hillman, R. E. (2003). “Measuring the frequency response function of laryngectomy patients: Implications for the design of electrolarynx devices,” Journal of the Acoustical Society of America **114** (2):1035-1047.

Moser, H. M., Oyer, H. J. (1958). “Relative intensities of sounds at various anatomical locations of the head and neck during phonation of the vowels” Journal of the Acoustical Society of America **30**(4):275-277.

Norton, R. L., and Bernstien R. S. (1993). “Improved Laboratory Prototype Electrolarynx (LAPEL): Using inverse filtering of the frequency response function of the human throat,” Annals of Biomedical Engineering **21:**163-174.

Paterkamp, H., Schäfer, J., Wodicka., G. (1996) “Posture-dependant change of tracheal sounds at standardized flows in patients with abstractive sleep apnea” Chest **110**:1493-1498

Shimamura, T., Tamiya, T. (2005). “A reconstruction filter for bone conduction speech” 48th Midwest Symposium on Circuits and Systems 2005, 1847-1850.

Snidecor, J. C., Rehman, I., Washburn, D. D. (1959). “Speech pickup by contact microphone at head and neck positions” Journal of Speech and Hearing Research 2(3):277-281.

Svec, J. G., Titze, I. R., Popolo, P. S. (2005). “Estimation of sound pressure levels of voiced speech from skin vibration of the neck,” Journal of the Acoustical Society of America 117:1386-1394.Welch, P.D. (1967), "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," IEEE Trans. Audio Electroacoustics, AU-15:70-73.

Wodicka, G. R., Shannon, D. C. (1990). “Transfer function of sound transmission in subglottal human respiratory system at low frequencies,” Journal of Applied Physiology **69**(6), 2126-2130.

Zhang, Z., Zicheng, L., Sinclair, M., Acero, A., Deng, L., Droppo, J., Huang, X., Zheng, Y. (2004). “Multi-sensory microphones for robust speech detection, enhancement and recognition,” Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). III p.781-784.

Zheng, Y., Liu, Z., Zhang, Z., Sinclair, M., Droppo, J., Deng, L., Acero, A., Huang, X., (2003). “Air- and bone-conductive integrated microphones for robust speech detection and enhancement,” IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) 2003, Virgin Islands, USA, 2003-11-30, p.249-254.