Frequency Response of the Neck during Production of Selected Speech Sounds

Jacob B. Munger, Scott L. Thomson
Department of Mechanical Engineering
Brigham Young University
Provo, UT, USA

Abstract. During speech the vocal folds vibrate resulting in audible sounds that are transmitted through the vocal tract as well as vibrations that are transmitted through the body tissue to the skin surface. These skin surface vibrations can be detected by contact microphones and used to transmit speech. The objective of this study was to characterize the frequency content of speech signals at a concentrated are on the neck. Signals were recorded using accelerometers attached to 12 locations on the neck of seven subjects as well as a microphone to record audible speech. The subjects produced several isolated phonemes. The power spectral densities (PSDs) of the phonemes were used to determine a quality ranking for each location and sound.

Introduction

During speech the vocal folds vibrate, resulting in audible sounds. In addition to being transmitted through the vocal tract, these vibrations are also transmitted through several layers of various types of tissue throughout the head and neck, resulting in small, but measurable, skin surface vibration. Contact microphones sense these skin surface vibrations for speech transmission, as opposed to acoustic microphones that sense air vibrations that radiate from the mouth.

Contact microphones have one significant advantage over acoustic microphones in environments with elevated ambient noise levels in that they sense very little background noise. In comparing the use of throat contact microphones to acoustic microphones for use in rotary-wing aircraft, Acker-Mills et al. (2004) found that throat microphones had approximately a 10 dB higher signal-to-noise ratio. Commercially available contact microphones, however, suffer from poor speech quality and intelligibility (Acker-Mills et al., 2004; Shimamura and Tamiya, 2005). This is a result of the skin vibrations being influenced by the many tissue layers (e.g. skin, fat, muscles, bones) of the neck or face between the contact microphone location and the vocal tract.

If the neck is the only option for microphone placement, it is desirable to locate it where the frequency response is the best. In this paper 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}) and the PSD_{SD} rankings.

Methods

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. 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 Institutional Review Board (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.

All accelerometers were manufactured by PCB Piezotronics (see Table 1 for specifications). All accelerometers were placed on the left side of the neck. 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. These accelerometers measure the magnitude and frequency of the skin vibration at each location while the subject speaks. An acoustic microphone was used to simultaneously acquire the audible speech.

The subjects sustained the vowels /a/ (b**a**t), /oo/ (b**oo**t), /ah/ (c**augh**t), /ee/ (f**ee**t), the nasals /m/ and /n/, and the fricative /f/ for 4 to 5 seconds each.

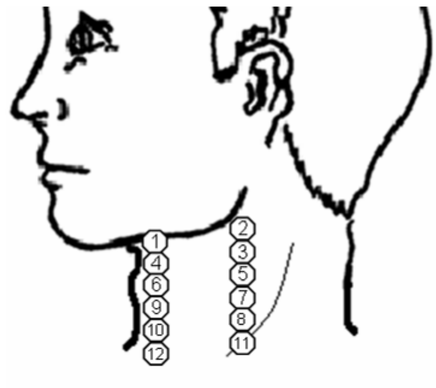

Figure 1 Accelerometer placement locations for neck location testing (image courtesy U.S. Army Research Lab Human Research & Engineering Directorate).

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

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

Data 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 was analyzed as follows. 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:

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

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 where 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 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}):

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

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 12 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 12 indicates the highest (worst) PSD_{SD} value.

Results

Power Spectral Density Summed Difference

Figures 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. A reduction in PSD_{SD} indicates a signal that better matches the microphone spectra. 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 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 2 and 3).

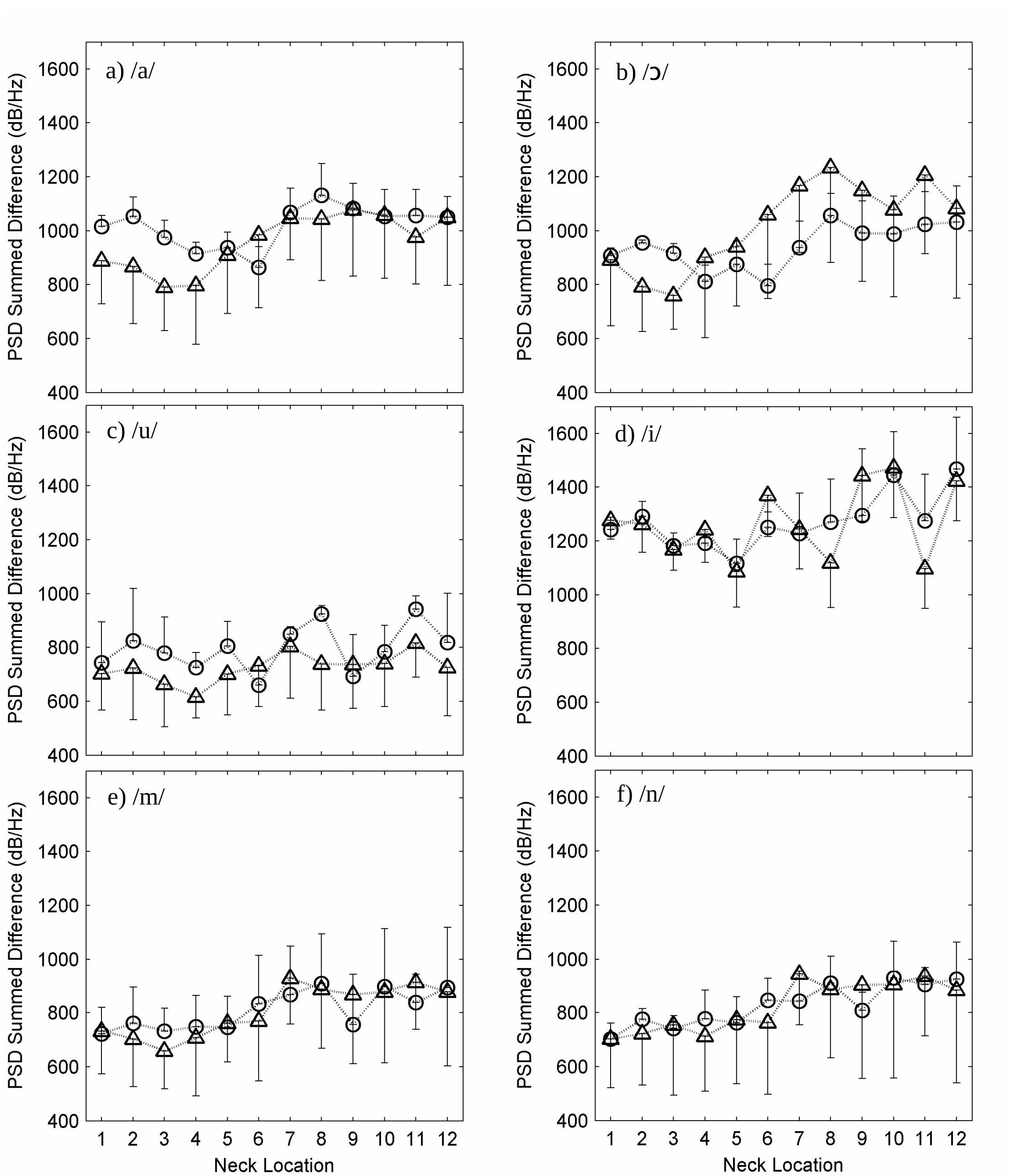

Figure 2 Normalized PSD_{SD} over 0-5 kHz for neck locations. ○: Male speakers; Δ: Female speakers. a) /a/; b) /ah/; c) /u/; d) /i/; e) /m/; f) /n/.

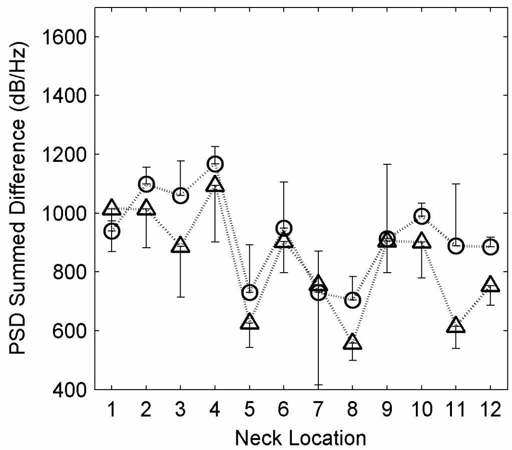

Figure 3 Sound /f/ normalized PSDSD for neck locations. ○: Male speakers; ∆: Female speakers.

Power Spectral Density Ranking

Tables 2 and 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 a 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 2 Average neck PSDSD rank, male speakers.

Neck	Sounds							
Location	/a/	/ah/	/i/	/u/	/m/	/n/	/f/	Average
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 3 Average neck PSDSD rank, female speakers.

Neck	Sounds							
Location	/a/	/ah/	/i/	/u/	/m/	/n/	/f/	Average
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

Discussion of Results

Influence of Location

Figure 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.

Figure 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. 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. 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.

Tables 2 and 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 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.

Gender Differences

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 indicates 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.

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.

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 /a/ and /ah/ 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 variation in the top ranked locations for the male subjects in this study it is recommended that further investigation 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.

Conclusions

In this paper 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.

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

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

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.