SPEECH LEVEL DISCREPENCY AROUND THE HEAD OF A TALKER – A NEW LOOK AT LAVALIER AND HANDSET MICROPHONE PLACEMENT TECHNIQUES

I.R Cushing Acoustics Research Centre, Salford University, Salford, U.K
F.F Li Acoustics Research Centre, Salford University, Salford, U.K
T.J Cox Acoustics Research Centre, Salford University, Salford, U.K

1 INTRODUCTION

Accurate speech signal acquisition is important in many applications and one important consideration is the best place to position near-field microphones. For precise speech intensity level and frequency analysis, discrepancies between speech signals recorded at different near-field microphones are of particular interest. The human voice has a complicated near-field pattern - sound propagation from the mouth is liable to be affected not just by diffraction and reflection effects of the environment but by the actual human body itself.

Many different applications depend on the reliable transmission of speech through near-field microphone positions, such as mobile phones, automatic speech recognition systems and television broadcasts. It is useful to compare such near-field measurements to that of a 1 metre reference microphone, as this is a typical speaker-listener distance in many conversational settings. This study aims to discover the level and frequency discrepancies between such microphone positions, in order to build compensation curves so that the level and frequency content of a 1m distance reference microphone can be matched to those measured on a microphone placed in the near-field. The result of these compensation curve filters will 'mimic' the speech signal as heard by a listener in a typical conversation. As previous work has shown, the acoustic-phonetic consequences of different near-field microphone positions[1] are often overlooked. This study will aim to take these discrepancies into account, using spectrographic analysis to look at some fine-structures of speech and the implications that such microphone differences may have on speech intelligibility.

2 METHOD

Recordings were made in a semi-anechoic chamber at the University of Salford to compare speech levels and frequencies between microphone positions. Three near-field microphones and one reference level microphone were used. They were positioned as follows:

(1) At the left cheek, as close to the corner of the mouth as possible.

(2) Clipped to the chest.

(3) Directly in front of the left ear.

(4) Reference microphone at a 1 metre distance from the speakers' mouth at the height of the speakers' ear.

Microphones 1 – 4 will be referred to as 'mouth', 'ear' 'chest' and '1m reference' respectively, from here onwards.

Omni-directional ½" GRAS 26CA microphones were used; all calibrated using an acoustic calibrator providing a sound pressure level of 94 dB re 20 µPa at 1000Hz. These high quality microphones are especially useful for precise speech measurements as they provide a flat frequency response across their entire range of 20 Hz to 20 kHz. 8 speakers were recorded in total; 6 males and 2

females. Speakers were seated and instructed to read a passage of text which was positioned directly in front of them. A 'comfortable' vocal effort level was stipulated. The phonetically balanced 'Rainbow passage' was used for the speech material, which took around 2 minutes for each speaker to read. The simultaneous recording of all microphones was made possible by using the NetdB multi-channel recorder, with the dBFA 4.8.1 software. Level and frequency content analysis were conducted using this software and MATLAB. Spectral analysis was performed via a narrow-band analyzer using a Hanning window.

3 RESULTS

Microphone	Males		Females	
	Average level dB(A)	Average frequency (Hz)	Average level dB(A)	Average frequency (Hz)
Mouth	78.0	117.8	74.7	191.3
Chest	72.2	116.8	71.6	191.4
Ear	67.5	115.5	62.1	191.8
Reference at 1m	55.3	121.0	52.6	195.9

Table 1: Average levels and frequencies

Table 1 shows the average level dB(A) and frequency (Hz) for speech collected at each microphone position, for the male and female groups. Average frequencies are calculated from 0 – 8 kHz. The upper frequency of the results is limited to 8 kHz as no significant speech information is present beyond this point.

On average, males were always louder than females, and females always higher in pitch. In previous work, Pearsons *et al*[2], carried out empirical research to acquire average speech levels in a variety of conversational settings. They found the average casual speaking level at 1m to be 52 dB(A) for males and 50 dB(A) for females. The results presented here are around 2 - 3 dB(A) louder, but the pattern for males being around 2 dB(A) louder than females remains the same.

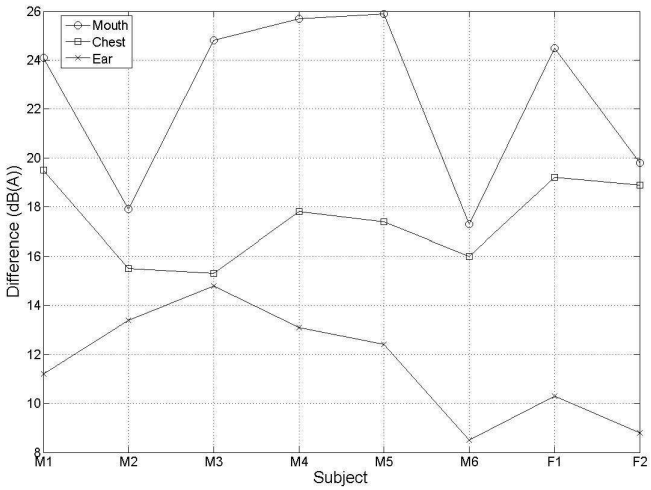

Figure 1: Individual subject level differences for the near-field microphones

Figure 1 shows the average level differences for each subject (M = males; F = females), for each near-field microphone position compared to the 1m reference microphone. The variation between individual subjects is rather high; for example M2 shows only a 4.5 dB(A) difference between the lowest and highest recorded levels. M5 on the other hand, shows a 13.5 dB(A) difference.

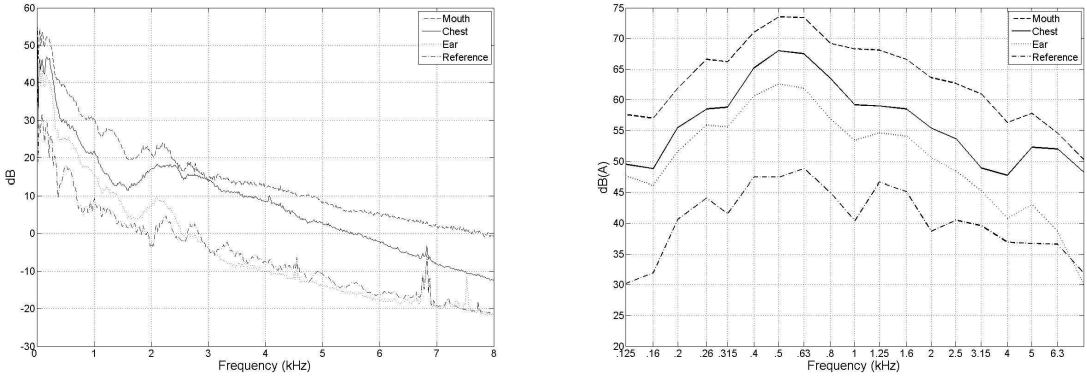

Figure 2: Spectra and 1/3 octave-band averaged for all subjects

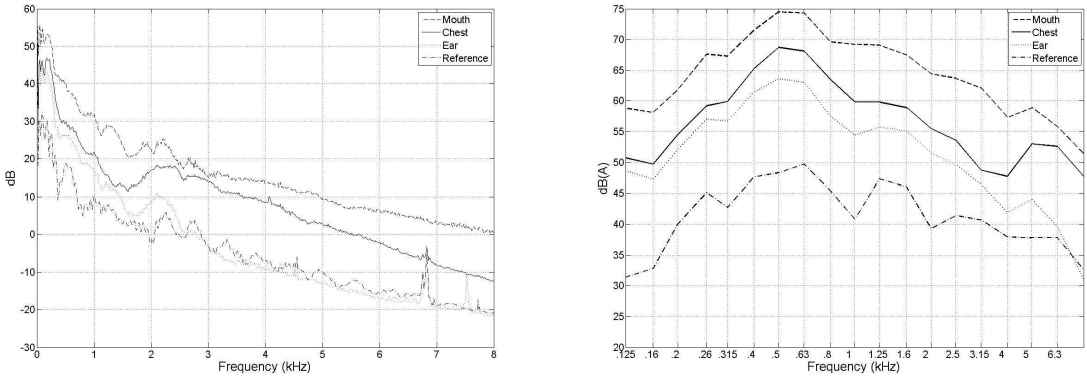

Figure 3: Spectra and 1/3 octave-band averaged for male subjects

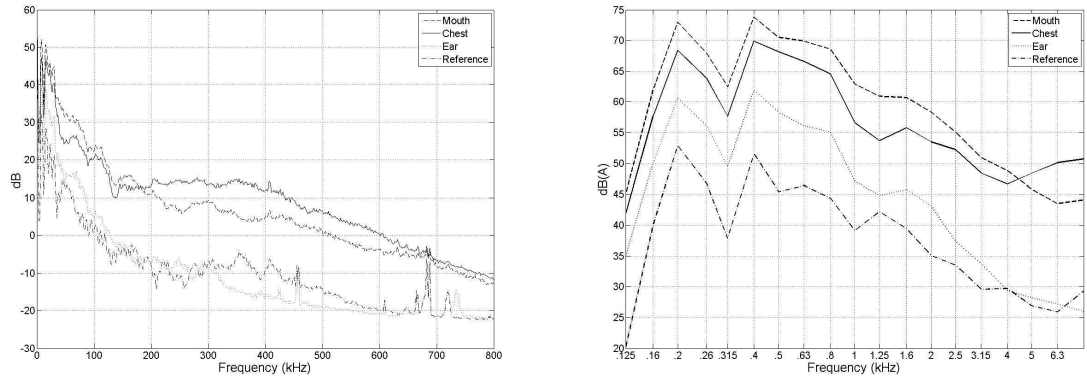

Figure 4: Spectra and 1/3 octave-band averaged for female subjects

Figures 2 - 4 show the spectra and 1/3 octave-band interval plots for each microphone position, for all subjects, the male subject group and the female subject group. All microphone positions have raised levels from 125 – 800 Hz, apart from a drop in dB(A) which centres on 315 Hz. Important speech information such as the fundamental frequency and certain vowels are contained within this region. The drop at 315Hz is especially noticeable in the 1/3 octave-band plot for the female subjects, where there is a dip of around 15 dB(A) from 200 – 315Hz. The drop in levels is much less dramatic for the male subjects and when averaged over all subjects, to only around 1 dB(A) (apart from the chest microphone which rises by 0.3dB(A)). Previous work such as that in Pearsons[2] and Brixen[4] reported attenuation at this frequency region, but not as dramatic as in the results gathered here.

From 630Hz, the general pattern seen is that levels decrease as frequency increases, but this reduction differs between microphones. From 0 – 4 kHz, the frequency responses for the near-field microphones are relatively similar; all curves follow the same pattern at different dB(A) levels, with a small amount of peaks and troughs. In the region of 4 – 8 kHz, frequency content differs much more between the near-field microphones.

3.1 Microphone at the mouth

Speech levels at this microphone are on average, 22.4 dB(A) above the level of the 1m reference microphone. Due to its position at the side of the head, there is a shadowing effect which results in the attenuation of high frequencies. This shadowing effect is not as extreme as seen at the ear (section 3.3).

3.2 Microphone on the chest

Speech levels at this microphone are on average, 18 dB(A) above the level of the 1m reference microphone. A dip in level is seen between around 3 – 5 kHz, due to the reflections of the human body. This has been well documented in previous literature[4].

3.3 Microphone at the ear

Speech levels at this microphone are on average, 10.9 dB(A) above the level of the 1m reference microphone. This matches the 1m reference microphone the closest, in terms of level differences. There is a notable dramatic drop in level at higher frequencies, due to a shadowing effect of the head. This results in levels dropping below even the 1m reference microphone at certain points.

4 COMPENSATION CURVES

To match the level and frequency response of a near-field microphone to that of a 1m reference microphone, compensation curves can be built based on the multi-channel recordings acquired for this study. The curves are computed by simply subtracting each near-field microphone away from the 1m reference microphone, at 1/3 octave-band intervals.

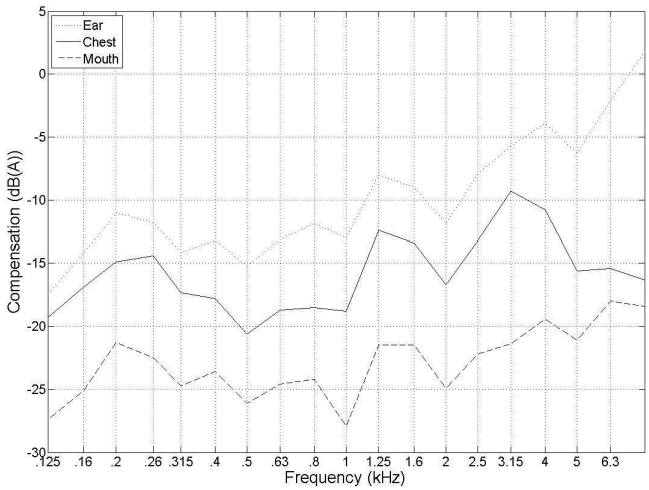

Figure 5: Compensation curves for each near-field microphone

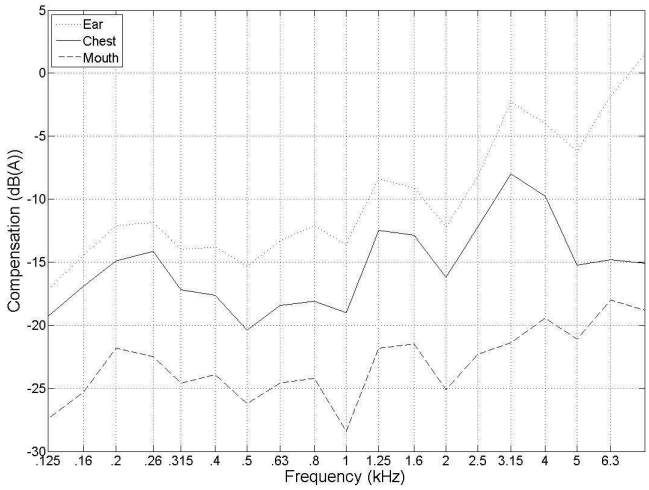

Figure 6: Compensation curves for each near-field microphone (for male speakers)

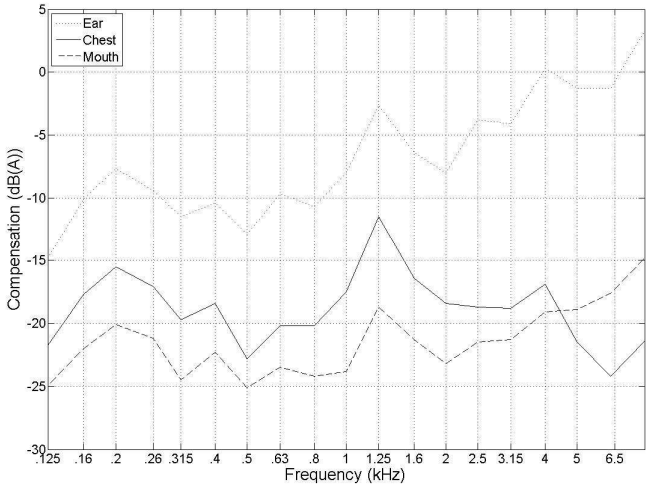

Figure 7: Compensation curves for each near-field microphone (for female speakers)

These curves will match the level and frequency response of each near-field microphone to a 1m reference microphone. They show the recommended dB(A) subtraction at 1/3 octave band intervals. It is relatively simple to implement these curves into a MATLAB function, to create automatic filters for the compensation curves.

5 ACOUSTIC-PHONETIC CONSIDERATIONS

The quality, viability and intelligibility of a speech signal can be degraded by a number of factors. Microphone placement is one of these, especially when speech is recorded in the near-field.

The frequency range of human speech is typically around 150 – 8000 Hz. British English vowel sounds fall from 150 – 2500 Hz, with consonants occupying frequency areas up to 8000 Hz. As shown in section 3, higher frequencies are more liable to be attenuated, meaning that the speech sounds most likely to be affected by near-field discrepancies are high frequency consonant sounds – fricatives and plosive burst onsets. It is hypothesised that these discrepancies will not have a negative effect on speech *intelligibility* but may deteriorate overall speech *quality*. Recording position discrepancies are evident in the spectrogram of speech shown in Figure 8:

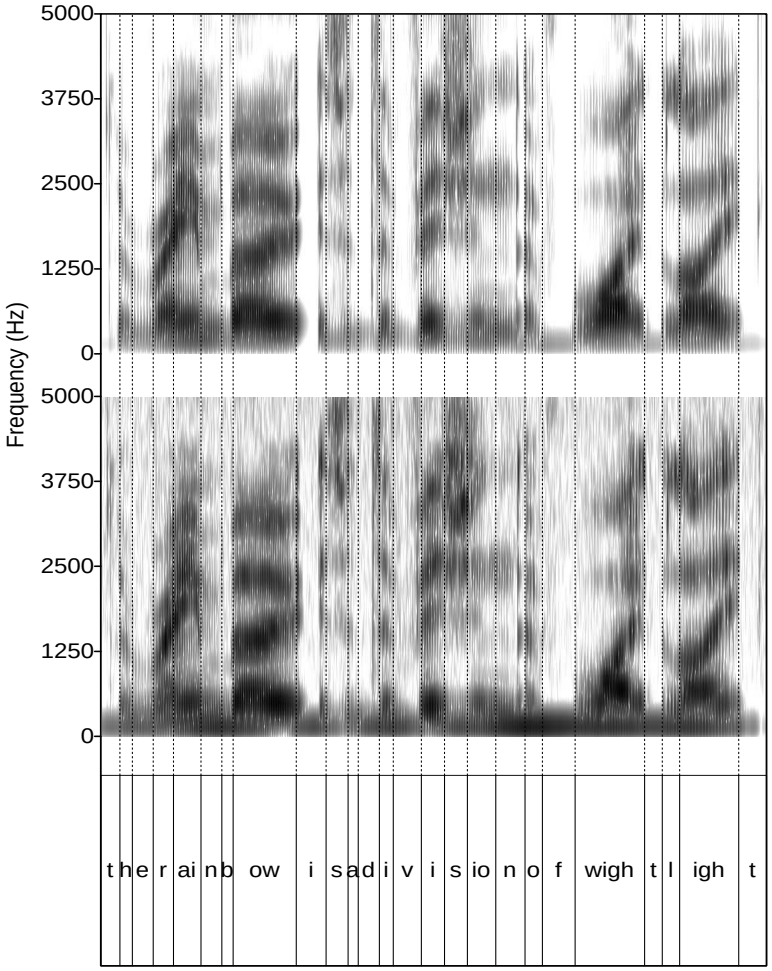

Figure 8: Spectrograms from the mouth (top) and 1m reference (bottom) microphones

Figure 8 shows spectrograms of the excerpt *the rainbow is a division of white light*, from a male speaker. The top spectrogram is from speech recorded at the mouth microphone. The bottom spectrogram is from the reference microphone.

Notice the brighter peaks and well defined transitions in the spectrogram from the mouth. The bottom spectrogram has more diffused energy and blurred transitions. This is especially evident on the final /t/ of 'light'; the burst is clearly evident on the mouth microphone spectrogram, but is hardly visible on the reference microphone spectrogram. It is likely that these differences will have implications for speech perception.

Brixen[4] hypothesised that different microphone positions could have serious implications for acoustic-phonetic analysis on a speech signal. With much current speech science research concentrating upon forensic phonetics, this concern is of particular importance and relevance. The effect of different microphone positions upon formant measurements was investigated by Hansen[1] and Plichta[5], who both found discrepancies across different microphone positions.

In order to conduct some preliminary acoustic-phonetic analysis, an /e/ vowel was taken from the recording of one male subject and its formant values compared across all microphone positions. The portion of the vowel analysed was identical in each case. The values were obtained using the automatic formant measurement algorithm in the Praat software.

Microphone	F1 (Hz)	F2 (Hz)	F3 (Hz)
Reference at 1m	453.9	1513.6	2275.9
Chest	509.6	1547.1	1843.6
Mouth	482.6	1634.1	2176.2
Ear	466.9	1628.7	1992.8

Table 2: /e/ vowel F1-3 values for one male subject at all microphone positions

Table 2 shows the F1, 2 and 3 values extracted from the /e/ vowel. Simply from this one example, it is clear that different microphone positions do have an effect on vowel formant measurements.

6 DISCUSSIONS AND CONCLUSIONS

The study revealed some interesting points to consider when choosing a suitable near-field microphone position for effective speech acquisition. Microphones placed at the mouth are least likely to be affected by reflections of the body and shadowing effects, due to their proximity to the sound source. However, they are prone to wind effects from the mouth which can lead to signal distortions and 'pops', especially on high energy consonant sounds such as /s/ and /p/. Another potential problem is that microphones at this position are not ideal if subtlety is a priority.

Speech levels were found to be 2 – 3 dB(A) louder than the *casual* levels given by Pearsons[2]. This could be for a number of reasons. One factor may be that in the results presented here, subjects were reading aloud as opposed to holding a conversation. Another likely reason for the increase in levels is the fact that these recordings were made in semi-anechoic conditions as opposed to anechoic conditions as made in 1977. Ground reflections from the floor will undoubtedly have some effect on the level differences.

As found in work by Halkosaari *et al*[3], attenuation of high frequencies is seen in microphone positions near to the cheek, a phenomena also reported in this paper. This attenuation increases the closer the microphone is to the ear, as evident in Figures 2 - 4. For the microphone placed at the ear, there is a sharp drop in level from 43.0 – 31.0 dB(A), from 5 – 8 kHz. A similar, but less sharp drop in level is seen at the mouth microphone, from 57.8 – 50.3 dB(A). The attenuation of high frequencies for microphones placed at the chest due to its reflective surface is well documented in the literature.

In terms of comfort for the user and inconspicuous placement, the chest mounted microphone is the clear choice. Microphones placed closer to the face may well make the wearer feel uncomfortable and have adverse effects on the speech. However, microphones worn on the chest may be prone to unwanted contact noise from clothing which can degrade sound clarity and quality. Different types of clothing may also have an influence on how much sound is absorbed at this position. These microphones also exhibit a dip in frequencies at 3 – 5 kHz.

The compensation curves given in section 4 will match the level and frequency response of any of the near-field microphones to the 1m reference microphone. This gives a typical talker-listener conversational effect; highly sought after in different speech applications. To the authors knowledge, this is the first study to construct compensation curves based on 1/3 octave band intervals.

Section 5 looked at the effect that microphone positions have on the accuracy of vowel formant estimation algorithms. From the preliminary analysis conducted and the results presented, it is evident just how much influence a different microphone position can have on extracting fine details of acoustic-phonetic elements. Similar work has been carried out by Hansen[1], and further work is required to quantify differences between other vowels and other parameters of speech that could well be affected. Other additional further work would involve applying the 1m compensation curves

to signals recorded at different microphone positions and carrying out subjective listening tests to see if people can perceive differences between them.

7 REFERENCES

1. Hansen, G & Pharao, N. (2006). Microphones and measurements. Lund University Working Papers In Linguistics, Vol. 52, 49-52.
2. Pearsons, K. S., Bennett, R.L., Fidell, S. (1977). Speech levels in various noise environments. EPA-6001-77-025, U.S Envrionmental Protection Agency.
3. Halkosaari, T. & Vaalgamaa, M. (2004). Radiation directivity of human and artificial speech. Workshop of Wideband Speech Quality in Terminals and Networks: Assessment and Prediction, Mainz, Germany.
4. Brixen, E. (1998). Near field registration of the human voice: Spectral changes due to position. Proceedings 104th Audio Engineering Society Convention.
5. Plichta, B. (2004). Data acquisition problems. In B. Plichta, *Signal acquisition and acoustic analysis of speech.*