FACE MOUNTED MICROPHONE PERFORMANCE

A Queenan* *Institute of Sound and Vibration Research, University of Southampton, UK*
*now at Hoare Lea Acoustics, Poole, UK

1 INTRODUCTION

Amplification of the voice using miniature microphones mounted on the face has been commonplace in theatre sound reinforcement for some time and it is widely believed in the theatre sound industry that the position of the microphone on the face affects the spectral content or, more generally, the "character" of the sound output. Sound practitioners strive to obtain the "best" sound, which introduces the concept of "better" sound, and there has been much experimentation but little controlled study of the variation of output signal with microphone position.

A four-stage undergraduate research project was carried out in an attempt to shed some light on this issue. In the first stage, some insight into industry knowledge and experience was obtained for this work through trade literature and direct communication with theatre sound practitioners. Comparisons were made between this information and physical analyses of the speech production processes in the second stage. In the third stage, the voices of seven talkers recorded using a number of microphones mounted on the face simultaneously were analysed for spectral differences. Finally, seven selected recordings of each talker were played to twelve listeners who were asked to compare each to a reference recording and note their observations and opinions in the fourth stage. The findings of the experimental stages were compared to those of the previous stages to determine correlation between industry experience and the laboratory results obtained.

2 BACKGROUND RESEARCH

2.1 THEATRE SOUND: INDUSTRY KNOWLEDGE AND EXPERIENCE

Despite their high cost both in purchase and rental, radio microphone systems have become a fundamental part of sound reinforcement for musical theatre around the world[1]. Combined with a concealed miniature electret microphone[2], possibly in one of a number of various skin tones, such systems provide what is demonstrably the best means of amplifying the voice with minimal visual intrusion (Figure 1). Certainly, practical alternatives involving PZM[3] or rifle microphones to pick up specific sources have been found to produce less satisfactory results.

Figure 1 - Miniature electret microphone mounted on the hairline

Discussions with a number of sound designers and operators have revealed some consistency in microphone placement preferences. The centre of the forehead is described as providing for the "clearest", "sharpest" and "nicest" sound and while the physical meaning of these qualities is unclear, it is clearly the preferred location for the practitioners in question. It is also noted from comments that the spectral content of the signal appears to change as the microphone is moved to the side of the forehead and down to the temple, with the extent and nature of this change being reportedly variable between performers. In some cases, the sound becomes "woolly" or "thin" and in others, there is a loss of clarity. Consensus regarding the sound at the cheek is again consistent; it has been described as "dull", "muffled", "boomy" and "muddy", with several respondents reporting loud bass and low-mid frequencies in the sound. One sound designer reports the level of the nose as the point beyond which the sound becomes undesirable.

2.2 SPEECH PRODUCTION MECHANISMS

Papanagiotou[4] describes a scheme for the classification of vocal sounds or phonemes which consists of three components; voicing, manner and place. Voiced sounds such as vowels are produced by the vibration of the vocal chords and are distinct from higher frequency unvoiced sounds, which are produced elsewhere in the vocal tract. The manner of articulation describes the mechanical action of the vocal tract and the place of articulation is the location of the main constriction. Under this scheme, the voice is modeled as the generation of sound at various points along the tract; a 'distributed source' (Figure 2).

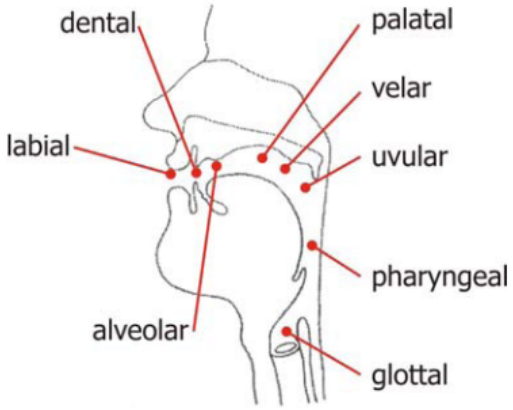

Figure 2 – Places of articulation (after O'Shaugnessy[5])

The tonal quality of the voice is defined in part by the resonance responses of air-filled cavities in the head, including the sinuses and oral and nasal cavities, to voice frequencies[6]. The volume of some cavities varies according to the manner and place of articulation; the oral cavity, for example, can be sealed almost entirely by labial, dental and alveolar constriction. While there is likely to be inter- and intra-subject variability in the resonance frequencies of these cavities, it also seems likely that the sound in nearfield of the head will be affected by them.

The results of multiple studies of the radiation of sound from a talker reveal clear back-to-front attenuation at all bands, corresponding with the ideal that the voice is loudest when the listener is inline with the notional "axis" of the talker[7,8,9,10,11,12,13,14,15]. The attenuation of sound increases with frequency as the listener moves off-axis around the head – the shadowing of high frequencies. In terms of directivity, A classic example of this is shown in Figure 3 in which the "axis" of the mouth appears to lie at 0º azimuth and between 0º and 45º elevation. It has been found that the pattern of radiation varies considerably according to the phoneme being produced, notably the size of the mouth aperture[9].

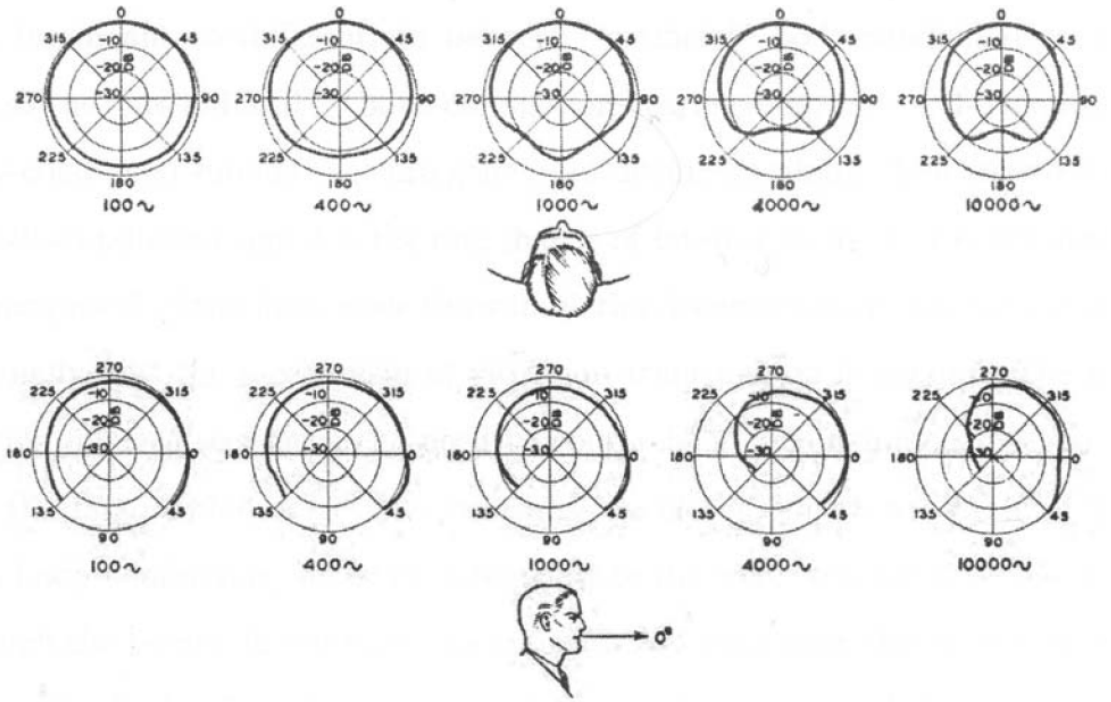

Figure 3 – Horizontal and vertical directional characteristics of the human voice

Previous study of the transmission of sound between the vocal chords and various locations around the head observed that the most faithful reproduction of vowels was recorded at the forehead, while the loudest sound however occurred in the laryngeal and mandible areas at the top of the neck[16,17]. The importance of the feedback loop mechanism which regulates the level and content of speech according to what the talker hears of himself was theorized and later proven[18].

Inter-subject variability in the quantities and proportions of bone, muscle and fat in the head will mean differences in the extent to which sound is conducted and absorbed. Genetic, behavioural and environmental factors may be compounded by age-related changes in the muscular and skeletal structures.

The distinction between spoken and sung vocalization has been the subject of study since the early 1970s and has advanced with the advent of superior analysis and modeling techniques. A 1972 study of vocal chord motion identified frequency-dependent phasing patterns during singing that were not discernible in the unvoiced sounds of speech[19], while a more recent study found that the control of the fundamental frequency of the voice is less tight in speaking than in singing[20]. Wide variation in pitch ranges and breathing patterns combine with issues such as training and vocal effort to produce a range of singing styles between ages, genders and musical genres. While the extent of the variation between singing and speaking voices is unclear, it seemed advisable to use spoken passages in this work for the purposes of this initial investigation.

3 EXPERIMENTAL RESEARCH

3.1 TEST TALKERS

The first group of experiments was intended to allow direct comparisons to be made between the spectral profiles of the voice recorded at various positions around the face. A passage of approximately phonetically-balanced text[Error! Bookmark not defined.] was read aloud by seven randomly selected test talkers, who were recorded using ten microphones mounted at various locations

around the face using a bespoke head-mount assembly. The same locations were used for each subject. A reference microphone was located approximately 1.5 metres in front of the talker.

The signals from all ten Monacor MCE-4000 electret microphones and the DPA 4060 reference microphone were amplified using ISVR-manufactured electret preamplifiers and acquired using an Alesis HD24 hard disk recorder. The frequency response of each of the Monacor microphones was expected to vary to some degree and clearly it was important to have some reference of this, so transfer functions for each microphone against the DPA reference microphone were obtained and were superimposed onto the spectra produced from the recordings. While the Monacor and DPA responses differed substantially with a maximum of approximately 9% at between 3kHz and 5kHz, it is interesting to note that the overall deviation between the Monacor microphones was found to be minimal (Figure 4).

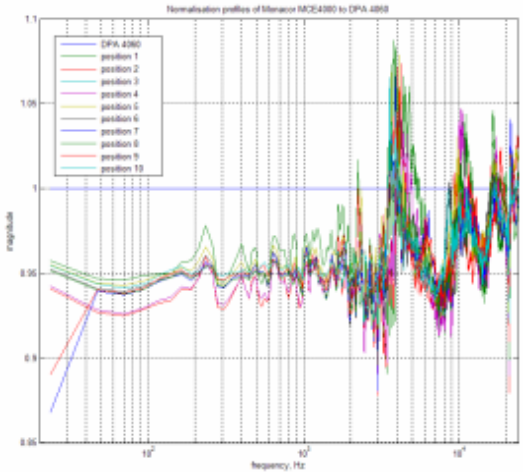

Figure 4 - Transfer functions of Monacor MCE-4000 microphones to the DPA 4060 reference microphone

Since the time available for analysis was limited in this work, long-term spectral averaging across the entire text passage was used; the average recording length was approximately 140 seconds. The shortcomings of this approach are discussed in section 4.

The measured long term averaged spectra for the microphones on each subject are shown in Figures 5 to 11. The signal level is highest at those positions closest to the mouth, i.e. the tip of the nose and the 1st maxilla position and the lowest levels are seen at different locations for different talkers. Figures 5 to 11 illustrate that the spectral differences seen are quite subtle in general, with none of the signals standing out as being greatly dissimilar to the others, implying that it might be possible to compensate for the differences using electronic equalisation*.

There is a region of distinct and often pronounced rippling at lower frequencies of the responses across all eleven microphones; this is particularly evident in talkers 3, 5 and 6 (females) and is minimal in talkers 4 and 7 (males). The effect is also seen in talkers 1 (male) and 2 (female), although there is a more pronounced drop in response between 100Hz and 110Hz in the latter. It is suspected that this is to do with the variation in distribution of fundamental voice frequencies between males and females. The typical male voice has an f_0 of between 85Hz and 155Hz (median average 130Hz) while that of the female voice is between 165Hz and 500Hz (median average 225Hz)[21]. It would be expected therefore that the spacing of vocal frequencies would be wider for females than for males who, in comparison, appear to have a smoother spectral profile across this

* Two sound designers have reported in response to this that, while this is true to an extent, a signal from a bad microphone position can never be made to sound as “good” as that from a preferred position. One goes on to describe the “good” and “bad” sounds as “natural” and “non-representative” respectively.

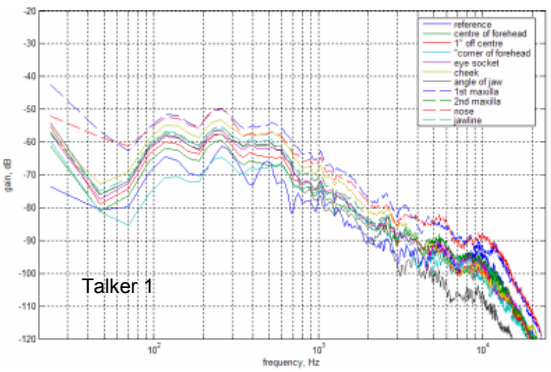

Figure 5 – Talker 1 averaged spectra

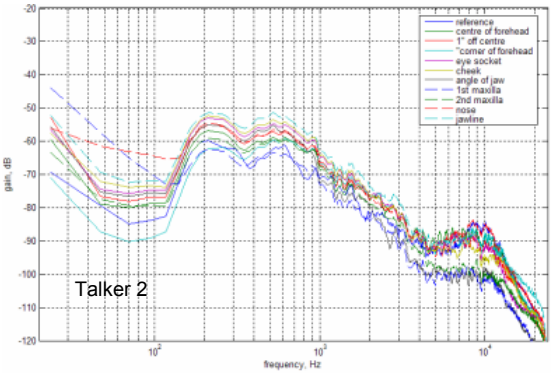

Figure 6 – Talker 2 averaged spectra

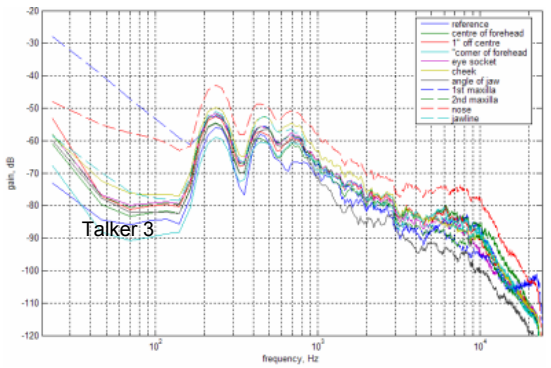

Figure 7 – Talker 3 averaged spectra

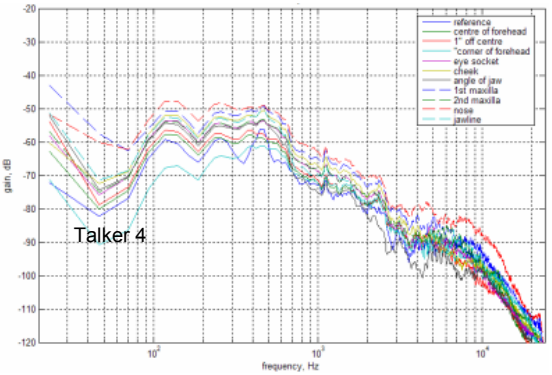

Figure 8 – Talker 4 averaged spectra

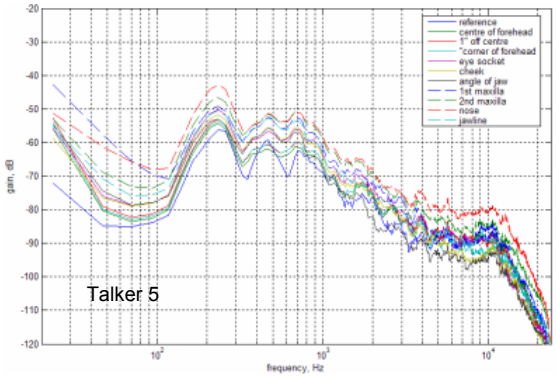

Figure 9 – Talker 5 averaged spectra

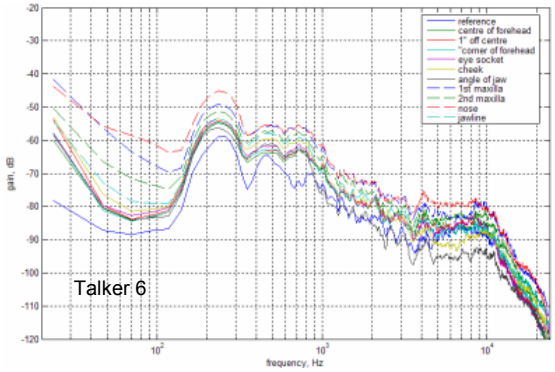

Figure 10 – Talker 6 averaged spectra

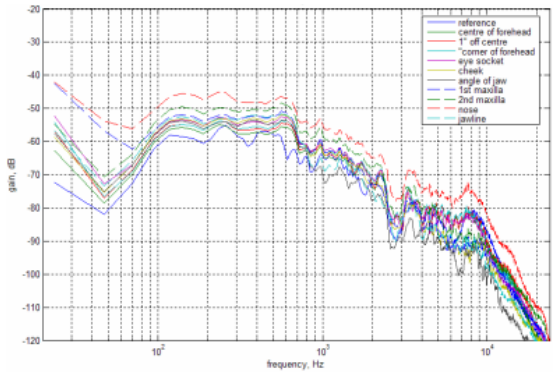

Figure 11 – Talker 7 averaged spectra

range. This is conducive with a lower f_0 and resulting higher density of frequencies and is corroborated by a distinct drop in output below the first and largest peak in talkers 2, 3, 5 and 6. Spectral content below these ranges may be considered artefact. The distinct peaks and troughs evident in all eleven signals (Figure 12) may correspond with formant frequencies, the most pronounced of which are likely to be those most commonly excited during speech, but the identification of specific formants falls beyond the scope of this work.

These areas may also indicate cavity resonances, although this seems less likely since there is little variation in shape between microphone positions. It is shown particularly clearly in Figure 13 that there is less high frequency energy at the cheek and angle of the jaw positions compared to the reference signal. It is unclear however whether this reflects a loss of high frequency content or boosting of low frequencies. There is elevation of between 5dB and 8dB at low frequencies at the cheek and it seems likely that this would be caused by cavity resonances.

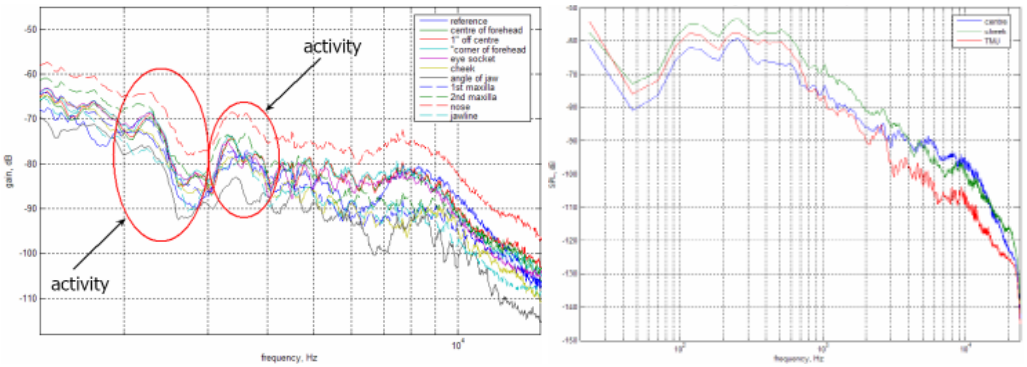

Figure 12 – Areas of activity that may be caused by voice formants or cavity resonances.
Figure 13 – Comparison of the spectra of signals from the centre of the forehead, the cheek and the angle of jaw

Figure 14 shows a comparison of the average levels seen at 300Hz and 10kHz compared to the reference signal. These frequencies appear to lie close to the lower and upper limits of the voice spectra shown in Figures 5 to 11 and may provide an indication of the proportion of high frequency content compared to low frequency content for each signal. The centre and corner of the forehead positions deviate least from the reference signal at these frequencies and the cheek, jaw line and angle of the jaw positions deviate most. Interestingly, the difference at the nose is greater than that at the forehead, which suggests that head shadowing may not be the most significant factor in this deviation.

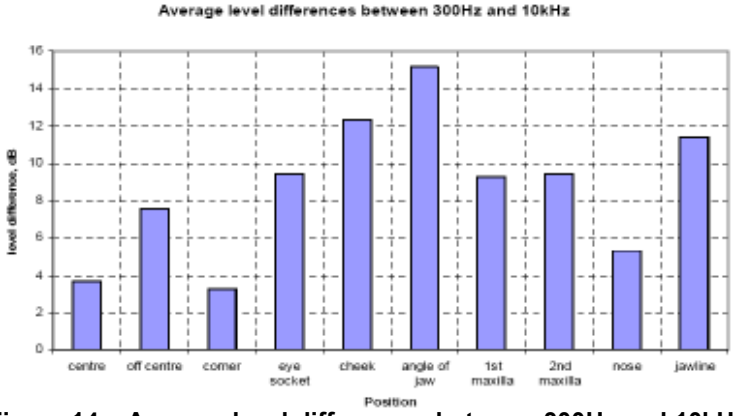

Figure 14 – Average level differences between 300Hz and 10kHz

3.2 TEST LISTENERS

The purpose of the second experimental stage was to obtain subjective impressions of the sound produced at each position. Twelve randomly selected test listeners were asked to compare the signals from seven[†] of the face-mounted microphones to the reference signal. To enable statistical analyses of the responses, listeners were presented with questionnaires in which they were asked to score each signal numerically using a -3 to +3 scale for four different qualities. It was explained to listeners that the scale was normalised to the reference signal, which would score zero for all four qualities. These qualities were described using descriptor pairs: dull versus bright, muffled versus clear, shallow versus rich and unpleasant versus pleasant. Listeners were also asked to make comments.

All eight signals were presented to each listener via a Behringer Truth active monitor and the listener was provided with a Yamaha O1X audio control surface in order to be able to listen to recordings individually in concurrent timing, or “solo” mode. The hard disk recorder was set to loop the whole passage and listeners were asked to take as long as they required completing the test. The seven face-mounted signals were normalised to the peak level of the reference signal, in order to prevent the possibility that level differences would colour observations of spectral differences.

Statistical analyses carried out suggest that distributions of perceptions are statistically significant for all four qualities. The sound at the centre of the forehead was rated as the most pleasant, with the off-centre and tip of nose positions almost as popular. The mean scores for all four qualities were greater than zero for the centre position but less than zero for the off-centre and nose locations, although the median scores for all four qualities were zero in all cases. This may imply that the sound at these positions is not necessarily better than the reference. The ANOVA calculations reveals that listeners are more sensitive to differences at the centre of forehead than they are at the cheek or jaw but less so to differences at the corner of the forehead. This suggests that it is harder to get a good sound at centre of forehead than to get a bad sound at the least popular locations. The largest proportion of negative responses was that seen at the angle of the jaw position. The sound was described as very muffled with a median score of -2 for dull/bright and this was also reflected in comments. These included “hollow”, “muffled” and “obstructed”, some or all of which may refer to spectral changes caused by head shadowing and/or cavity resonances.

Breath noise was mentioned in half of all comments on recordings at the tip of the nose and is clearly audible and unmistakable. However, the second most popular comment for tip of nose was that it was better than reference, with two others reported that it was the same. It is likely that breath noise obscured an otherwise high quality result at the tip of the nose. This seems intuitive since the

[†] It was not possible to replay all ten face mounted signals due to equipment limitations.

location is very close to the mouth, meaning a lesser contribution from head shadowing there than at any other position.

The response peaks for each of the four qualities in Figures 15 to 21 are clustered around the same values in many cases. This may indicate interactions between qualities, some degree of ambiguity in the meaning of descriptors pairs or that there is actually not much difference between the subjective quality of the recordings. Two listeners mentioned that they did know what was meant by "shallow" or "rich". The median score for this quality was zero in five out of seven cases and the ANOVA F-ratio was lower, so it is possible that the significance of results for this quality may be limited. It is also interesting to note that very few listeners awarded high and low scores for different descriptors for a single recording. This may again reflect ambiguity in the meaning of descriptors but it is also possible that there is a generalized prejudice in which "nice" sounds are awarded high scores in all qualities and, conversely, that less "nice" signals receive low scores.

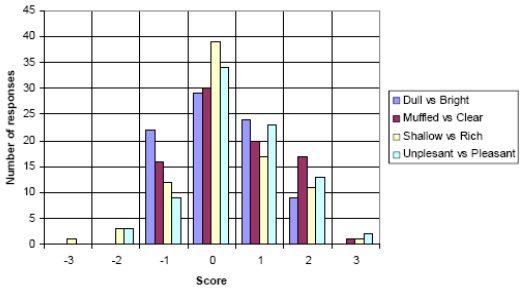

Figure 15 – Centre of the forehead

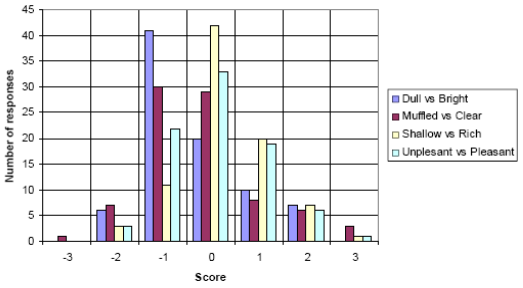

Figure 16 – One inch off-centre

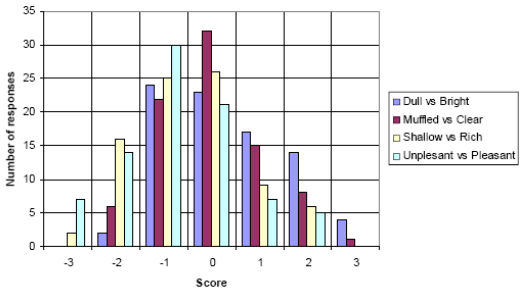

Figure 17 – Corner of the forehead

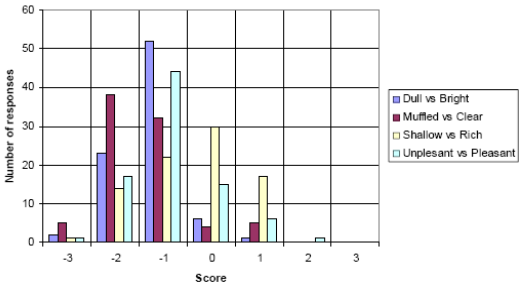

Figure 18 – Cheek

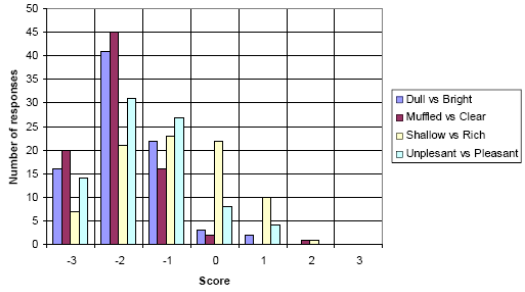

Figure 19 – Angle of jaw

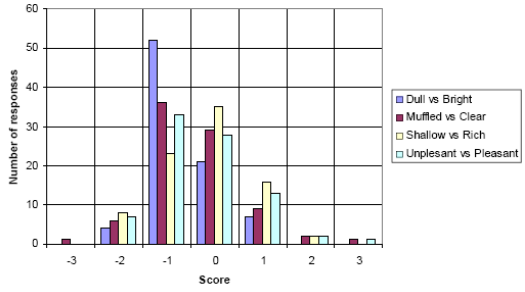

Figure 20 – 2nd Maxilla position

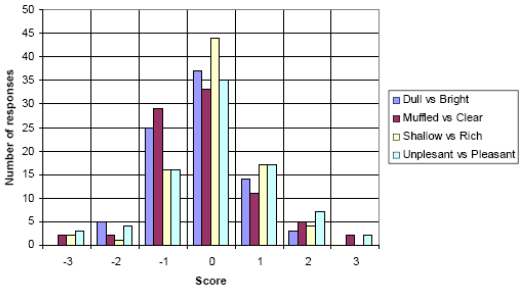

Figure 21 – Tip of nose

The variances of results for all four qualities are largest at the corner of the forehead which suggests either that there is less agreement between listeners at this position or that the sound is less consistent between talkers. Some comments describe it as “radio-like” and “phased”. However, the highest and second-highest scores for dull/bright and muffled/clear respectively are seen at this position, which may therefore be the best location to use where optimum intelligibility and clarity, rather than pleasantness, are required.

Interaction between talker gender and listener scores has been investigated statistically. There does not appear to be any significance overall, except at the tip of the nose where there is strong interaction. The meaning of this is unclear but may be to do with the length of the nose with respect to the cone of directivity of the mouth or, alternatively, differences in the volume or geometry of the nasal cavity between genders. Two interesting comments were made reporting that it was easier to score the male voices than the female voices, with one listener reporting that there was “somehow more to the male voices”.

4 LIMITATIONS AND FURTHER WORK

The exploratory nature of this work, which was carried out as an undergraduate project over a period of seven months, means that a number of areas were identified in which further investigation would have been useful had time been available. For example, the acquisition of more data by using a larger number of talkers and listeners is an obvious progression from this work.

The long-term averaging used in this work imposes restrictions upon the identification of anything other the average spectral content in measurement data. Spectrograms produced from the recordings would be expected to reveal voice formant activity, from which it may be possible to determine correlation between spectral content in general and the subjective responses acquired.

Subjective differences in vocal tract length and articulation dynamics mean that the identification of specific resonances may require the development of techniques that would make it possible to measure cavity volumes. This may lead to the development of theory regarding the impact of each resonance upon the sound produced at different microphone positions.

Talkers and listeners were not asked if they had received speaking, singing or listening training of any kind prior to participating in this work. It is assumed that all participants were untrained and this assumption may be invalid. It is possible that a correlation exists between the manner and extent of training in both talkers and listeners and the results of both the objective spectral and subjective opinion experiments.

5 CONCLUSIONS

Substantial insight into the impact of microphone position on the perceived sound quality has been obtained through this work. There is clear interaction between industry experience, speech

production mechanisms and the results of the talking and listening tests. There is clear agreement that the centre of the forehead provides for the subjectively best sound while the sound at the cheek and angle of jaw are the least popular. Attenuation at high frequencies, suspected to be due to head shadowing and spectral differences resulting from cavity resonances provide the most compelling explanation for these observations. Patterns in listener responses to the recordings presented appear to be statistically significant while the gender of the talker does not.

6 REFERENCES

[1] Deans, J. (2000). Tricks of the Trade: Using Lavalier Microphones. Mix Magazine.
[2] Electret Microphone Turns 40, Lucent Technologies. (2002).
[3] Burroughs, L. (1974). "Microphones: Design and Application". Sagamore Publishing Company, Plainview, New York.
[4] Papanagiotou, K. (2003). Enhancement of Body-Conducted Speech from an Ear-Microphone. PhD thesis. ISVR, University of Southampton.
[5] O'Shaugnessy, D. (2000). "Speech Communications: Human and Machine." IEEE Press.
[6] Norman, W. (1999). Skull, Scalp and Superficial Face: Head and Neck. URL: http://mywebpages.comcast.net/wnor/lesson1.htm
[7] Dunn, H., Farnsworth, D. (1939). Exploration of pressure field around the human head during speech. J. Acoust. Soc. Am. (10) pp.184-199
[8] Kuttruff, H. (1979). Room Acoustics. Elsevier Applied Sciences (Publishers) Ltd, New York.
[9] Halkosaari, T., Vaalgamaa, M. (2004). Directivity of human and artificial speech. Workshop of Wideband Speech Quality in Terminals and Networks: Assessment and Prediction, Mainz, Germany.
[10] McKendree, F. (1986). Directivity indices of human talkers in English speech. Inter-Noise, MIT, Cambridge, MA.
[11] Huopaniemi, J., Kettunen, K., Rahkonen, J. (1999). Measurement and modelling techniques for directional sound radiation from the mouth. IEEE Workshop in Applications of Signal Processing to Audio and Acoustics, New York.
[12] Olson, H. (1967). Music, Physics and Engineering. Dover (Publishers) Inc., New York.
[13] Kob, M., Jers, H. (1999). Directivity measurement of a singer. ACTA Acustica 85(S1).
[14] Flanagan, J. (1960). Analog measurement of sound radiation from the mouth. J. Acoust. Soc. Am. 32: pp. 1613-1620.
[15] Studebaker, G. A. (1983). Directivity of the human vocal source. J. Acoust. Soc. Am. 73(1): S105.
[16] 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. J. Acoust. Soc. Am. 30(4): pp. 275-277.
[17] Bekesy, G. (1949). The structure of the middle ear and the hearing of one's own voice by bone conduction. J. Acoust. Soc. Am. 21: pp. 217-232.
[18] Maurer, D., Landis, T. (1990). Role of body-conduction in the self-perception of speech. Folia Phoniatrica 42: pp. 226-229.
[19] Titze, I. R., Strong, W. J. (1972). Simulated vocal chord motions in speech and singing. 83rd Meeting of the Acoustical Society of America, Buffalo, NY.
[20] Chou, W., Gu, L. (2001). Robust singing detection in speech/music discriminator design. IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT.
[21] Baken, R. J. (1987). Clinical Measurement of Speech and Voice. Taylor and Francis (Publishers) Ltd, London.