Proceedings of The Institute of Acoustics

THE EFFECT OF AN OXYGEN MASK ON THE SPECTRUM OF SPEECH

S J Elliott and P Darlington
Institute of Sound and Vibration Research
University of Southampton

Introduction

When speech is uttered into a mask, in this case an oxygen mask of the type used in high performance aircraft, it is distorted in a number of ways. One obvious distortion is that the average spectrum is completely different from that of speech uttered under normal conditions. This paper reports an attempt to measure this spectral change and to begin to explain the mechanisms which cause it in terms of the acoustics of the speech production process. Since the frequencies of interest are low enough that sound propagation in the vocal tract is essentially plane, we can represent the speech source at the mouth with a simple one port electrical analogue. This in turn can be completely described by its Thevenin equivalent; a perfect pressure source (P_t) in series with an acoustic impedance (Z_t). Previous work in which this analogy was used [Ref 1], assumed that the vocal source, at the lips, had constant volume velocity, i.e. that Z_t was considerably larger than both the radiation impedance (Z_r) seen by the lips when speech is uttered under normal conditions (producing a pressure P_r, say), and the acoustic impedance of the mask as seen at the lips (Z_m) when speech is uttered into the mask, producing a pressure P_m. Under these circumstances the ratio of the two pressures (P_m/P_r) is equal to the ratio of the two impedances (Z_m/Z_r). During a series of initial experiments the acoustic impedances looking into the mouth (Z_t), into a mask (Z_m) and due to radiation at the lips (Z_r) were measured. It was found that Z_m was comparable with Z_t and became greater than Z_t over certain frequency ranges. This casts considerable doubt on the assumption that the vocal source at the lips has a constant volume velocity. The Thevenin equivalent can still be used, however, to derive the ratio of the two pressures;

$$\frac{P_m}{P_r} = \frac{Z_m}{Z_r} \left[\frac{Z_t + Z_r}{Z_t + Z_m}\right]$$

Z_t, however, depends on the shape of the vocal tract, i.e. what is being said. Thus this ratio would, at best, be consistant only for a steady utterance. The approach taken at this stage was to directly estimate the ratio P_m/P_r as a function of frequency for various isolated utterances and connected speech with a number of subjects.

Measurement of the spectrum of speech in the free field and in a mask

The pressure waveform of speech when spoken in a free field room was picked up using an instrumentation microphone (0.5" Bruel and Kjaer type 4134) placed a fixed distance (approximately 0.4m) from the subject's mouth. The averaged power spectral density was then calculated using a Solatron 1200 frequency analyser with a frequency range of 0 to 5kHz, and averaging over 32 individual spectra (about 3 seconds of speech). Similarly the pressure waveform of speech in the mask was picked up with the same instrumentation microphone sealed into the hole in the mask into which the communications microphone is normally placed. The microphone when placed in the mask had a thin (1/8") layer of open pored foam placed across it to reduce breath noise.

The averaged spectrum was measured, for a given subject in the free field room and in the mask, for various segments of speech. These consisted of:

a) Isolated continuous vowels and consonants;-
"EE", "AW", "MM", and "SS" and

b) An example of connected speech (the phrase "Joe took fathers shoe bench out, he was waiting at my lawn" repeated twice).

Five male subjects were used with ages ranging approximately from 20 to 30, who were instructed to speak at a "normal" conversational level in both the free field and mask conditions.

For the voiced examples of isolated speech the spectrum approximates a line spectrum, with appreciable energy only at harmonics of the fundamental. When a subject utters this sound on different occasions, for example in the free field and then into the mask, there are inevitable small differences in the fundamental frequency, which would make it very dubious to just subtract the two levels of the spectra to obtain the level of the ratio of the two. One method of avoiding this would be to numerically estimate the envelope of the line spectra in each case and subtract these. However, it was found that if the pitch of the utterance was slowly varied up and down by about a third of an octave ("warbled"), the resulting average spectrum was a good approximation to the envelope of the line spectra produced with constant pitch.

For a given subject and utterance the measurements of the spectra with or without the mask were repeated three times and the spectral densities found to be repeatable to within ±3dB on each occasion. This repeatability gave some confidence in the subsequent operation of subtracting the spectral density levels measured for a given subject and utterance with the mask, from those without. The resultant curve is the logarithm of the ratio of the pressure measured in the mask to that measured in the free field, as a function of frequency and with units of dB.

Notice that because the same microphone has been used for both measurements then the microphone respopnse does not effect the difference, although because of signal to noise considerations it is best to use a microphone with a flat response such as the one used. Another advantage of using the same microphone is that calibrated differences in level are produced without having to worry about the sensitivity of the microphone.

The spectral difference results show a certain amount of fine variation with a distinct trend superimposed. The fine variation is because neither of the individual spectra are completely smooth, even with pitch variation, especially at the lower frequency end. The overall trends in the curves are, however, found to depend on both the utterance and the speaker.

Due to lack of space we are unable to reproduce all the results, a representative selection are shown in Figure 1 for three utterances and two subjects.

Conclusions

The results show a general lift at low frequencies of about 50dB falling to about 30dB by 1kHz. The average trend across utterances is reasonably flat above this frequency as shown by the connected speech results. However for the isolated segments of speech there are seen to be dips and peaks superimposed on the overall trend, which may be of the order of 20dB amplitude, and whose frequency and width are both utterance and speaker dependent.

Whereas it is to be expected that these curves would look different for various utterances because the vocal tract impedance (Z_t) depends on the way each utterenace is articulated, it is at first sight surprising that there is so much variation between subjects. Possible reasons for this are that either each subject produces a perceptually similar sound with a slightly different vocal tract configuration and hence produces a different Z_t, or that the mask impedance (Z_m) is different from subject to subject because of face shape and/or fitting. Both effects may act together to produce the observed results.

Acknowledgements

This paper arose out of work undertaken by the Auditory Communication and Hearing Conservation Unit of the ISVR, managed by Mr P Wheeler, supported by the Ministry of Defence, Procurement Executive.

Reference

E. Singer. "The effects of microphones and face masks on LPC vocoder performance". Int. Conf. Speech and Signal Processing (1981), 3, pp.1066-1069.

FIG. 1 RATIO of PRESSURE SPECTRA in MASK and in FREE FIELD.