SPEECH ENHANCEMENT IN NOISE AND WITHIN FACE MASK
(Microphone Array Approach)

George S. Kang and *Thomas M. Moran*

Naval Research Laboratory
Washington, DC 20375
kang *or* moran @itd.nrl.navy.mil

ABSTRACT

In certain communication environments, digital speech transmission systems must work in severe acoustic environments where the noise levels exceeds 110 dB. In other environments, speakers must use an oxygen face mask. In both situations, the intelligibility of encoded speech falls below an acceptable level. We have developed a technique for improving speech quality in these situations.

Previous speech improvement methods have focused on processing the corrupted signal after it has been induced by the microphone. These methods have not performed adequately. In our technique, speech anomalies are attenuated by a microphone array before speech and noise become mixed into a signal.

Our microphone array prototype has shown excellent performance. In an example of speech taken aboard an E2C aircraft, this noise-canceling microphone array improved the speech-to-noise ratio by as much as 18 dB. When the same technique is used in a face mask, muffled speech was almost completely restored to high quality speech.

1. INTRODUCTION

Over the years, various signal processing techniques have been developed to reduce background noise in a speech signal [4]. It is difficult, however, to improve speech by these noise reduction techniques once speech and noise have become mixed together. In fact, a study by the National Research Council concluded that noise reduction processing rarely improved speech intelligibility [3]. Therefore, we present an approach in which noise is reduced at the microphone before it can become mixed with the speech signal. We use a microphone array to accomplish this objective.

At the outset, we would like to state that we do not use the type of microphone array which forms a highly directional microphone gain pattern. That type of array focuses on the speech source, while rejecting noise sources arriving from other directions. It has been widely studied in recent years for use in hands-free cellular phones in automobiles and other similar applications. Unfortunately, in this type of microphone array, both the mouth-to-microphone distance and microphone element spacing are measured in feet, making it too large to be placed in handsets or face masks.

Our microphone array, rather than being highly directional, achieves noise reduction by having a large attenuation per source distance. In other words, the array mostly picks up nearby sounds while rejecting distant sounds. Since the mouth is located 1/4 inch away from the array, and noise is usually coming from further away, the array picks up more speech and less noise. This characteristic is obtained by taking the difference between the signals coming from adjacent microphone elements, at successive stages. In this type of microphone array, more attenuation per source distance results if the microphone elements are placed closer together (a fraction of an inch). Therefore, such a microphone array can be placed in a handset or within a face mask.

Microphone arrays of this type were originally investigated by Harry Olson in the 1940s [5]. The first attempt to implement this type of microphone array was made by Beaverson and Wiggins in 1950 [1]. They used a microphone containing a single acoustic diaphragm having four acoustic paths. The effort was not too successful because source attenuation at a distance of one inch was only 3.5 dB (rather than the 20 to 30 dB necessary for noise reduction). In 1981, a similar effort was also reported, again using a single diaphragm [2].

Our array has four separate microphone elements. Contained within each element cartridge is an electret microphone and a high-impedance field-effect transistor. Each cartridge is about 6 mm in diameter by 3 mm in height. Matching microphone characteristics is not a serious problem when using an electret microphone element, because they have a virtually flat frequency response over the frequency range of interest. The gain of each element was individually controlled in order to obtain a matched amplitude response.

2. MICROPHONE ARRAY

A microphone outputs a voltage proportional to the sound pressure at the microphone diaphragm [5]. Thus,

$$e = A\frac{\sin\frac{2\pi}{\lambda}(ct - r)}{r}, \tag{1}$$

where c is the speed of sound, λ is wavelength, r is source distance, t is time, and A is a factor related to sound intensity converted to electric amplitude.

Let the four microphone elements be located in symmetric positions at $[x(j), y(j)]$, j = 1, 2, 3, and 4 (see Fig. 1), where the center of the array corresponds to the origin of the coordinate system. Let the sound source be located at a radial distance R with an angle of θ. Then, the distance between the sound source and the jth microphone elements, denoted by $r(j)$, is expressed by

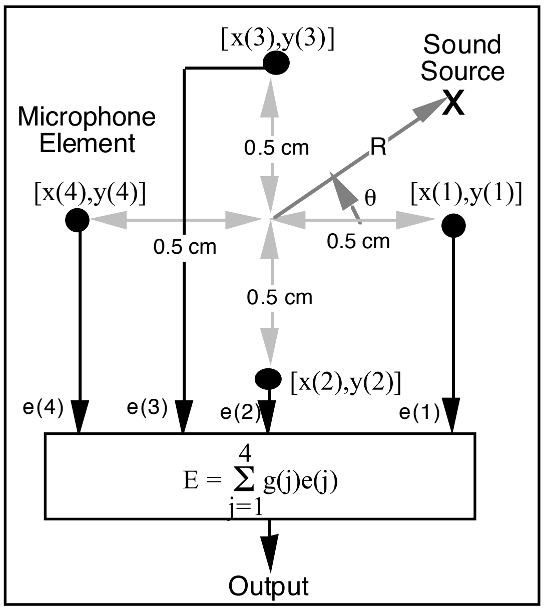


Fig. 1 — Generalized four-element microphone array configuration. A collinear array (shown in Fig. 3) is a special case of the array shown in this figure.

$$r(j) = \sqrt{\left[x(j) - R\cos\theta\right]^2 + \left[y(j) - R\sin\theta\right]^2}, \qquad (2)$$

where j = 1, 2, 3, 4. From Eqs. (1) and (2), the induced voltage in each microphone element is expressed by

$$\begin{aligned} e(t,\lambda,j) &= A\frac{\sin\frac{2\pi}{\lambda}\left[ct - r(j)\right]}{r(j)} \\ &= A\frac{\sin\left[\omega t - \frac{2\pi}{\lambda}r(j)\right]}{r(j)} \end{aligned} \qquad (3)$$

where $\omega = 2\pi c/\lambda$ is the source frequency in rad/s. The output of the array, is expressed by

$$E(t,\lambda) = A\sum_{j=1}^{4} g(j)\left[\frac{\sin\left[\omega t - \frac{2\pi}{\lambda}r(j)\right]}{r(j)}\right]$$

or,

$$\begin{aligned} E(t,\lambda) = {} & A\sum_{j=1}^{4} g(j)\frac{\cos\left[\frac{2\pi}{\lambda}r(j)\right]}{r(j)}\sin\omega t \\ & - A\sum_{j=1}^{4} g(j)\frac{\sin\left[\frac{2\pi}{\lambda}r(j)\right]}{r(j)}\cos\omega t \end{aligned} \qquad (4)$$

where g(j) is the gain for the jth microphone element. From Eq. (4), the magnitude of the array output is expressed by

$$\left|E(t,\lambda)\right| = A\sqrt{\left(\sum_{j=1}^{4} g(j)\frac{\cos\psi}{r(j)}\right)^2 + \left(\sum_{j=1}^{4} g(j)\frac{\sin\psi}{r(j)}\right)^2} \qquad (5)$$

where ψ is a short-hand notation for $(2\pi/\lambda)r(j)$, and the ideal gains are $g(1) = g(4) = 1$ and $g(2) = g(3) = -1$ in the absence of gain mismatch.

Computed Frequency Response

The frequency response at various source distances is an important property of a noise-canceling microphone. We plotted the theoretical amplitude response of the microphone array by using Eq. (5). The result is shown in Fig. 2.

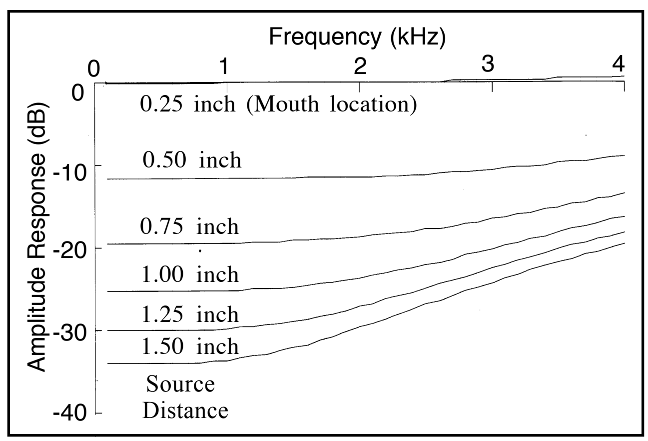


Fig. 2 — Theoretical frequency response of the microphone array in which four microphone elements are placed collinearly 0.5 cm apart. Eq. (5) is used to plot these curves.

Note that the near-field response (for speech) is virtually flat, and that the far-field response (for noise), though not flat, exhibits a large amount of attenuation. For example, attenuation at one inch from the speech source (speaker's mouth) at a frequency of 1 kHz is almost 30 dB. It is this large attenuation per source distance that enables the array to remove noise. As noted in Fig. 2, noise reduction is less effective for higher source frequencies. This is not a significant drawback if background noise contains predominately low frequencies (such as vehicular engine noise, which is our concern).

3. PROTOTYPE

As previously stated, we use four electret microphones, each with a high-impedance field-effect transistor packaged into a cylindrical cartridge (about 6 mm in diameter by 3 mm in height). These microphone cartridges are spaced collinearly by placing them 0.5 cm apart in a rectangular block which we prepared (Fig 3).

First, the sound duct is made by boring a cylinder 6 mm in diameter through the length of the block. Then a series of four microphone element shafts are drilled perpendicular to the sound duct at intervals of 0.5 cm. Because the spacing interval of these shafts is less than their diameters, adjacent shafts must be drilled from different sides of the block (Fig. 3). For the microphone housing material, we tested both foam rubber and wood. Results are similar.

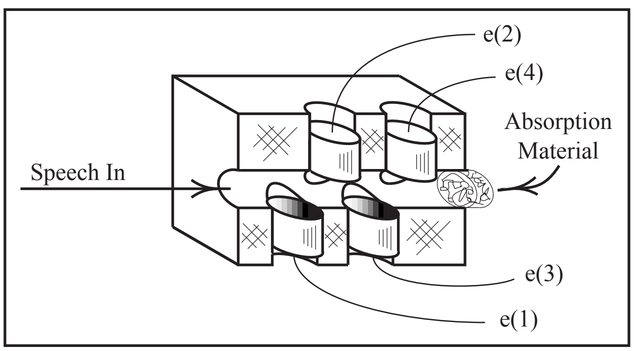


Fig. 3 - Cross-sectional view of the microphone array housing used in our prototype where four microphone elements are spaced collinearly at 0.5 cm apart.

Some tuning of the sound duct is necessary to reduce unwanted resonance. This was done in a trial-by-error method by stuffing the sound duct with various sound absorbing materials until a flat frequency response was achieved. Note that the absorption coefficient of each material is frequency-selective [6]. For example, wool absorbs sound energies well for frequencies around 4 kHz.

The amplifier performs the operation [e(1)+e(4)]-[e(2)+e(3)], where e(1) through e(4) are individual microphone outputs. Because the difference between [e(1)+e(4)]-[e(2)+e(3)] is small, the amplifier must be designed and fabricated with care. The quality of the summing amplifiers critically affects the noise reduction performance of the microphone array.

Measured Frequency Response

Figure 4 is the actual amplitude response of the microphone array prototype measured by using a Brüel & Kjaer Audio Analyzer. The attenuation at 1 inch away from the mouth location is as much as 24 dB. This is 10 to 15 dB greater than that of existing noise-canceling microphones. In addition, the array has no sharp resonant frequencies which are often found in existing noise-canceling microphones.

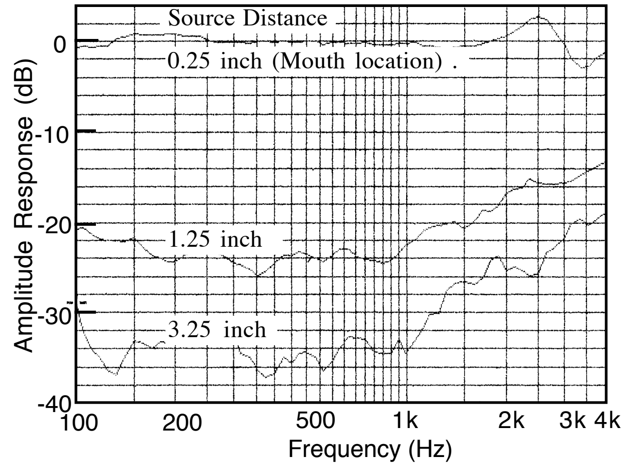
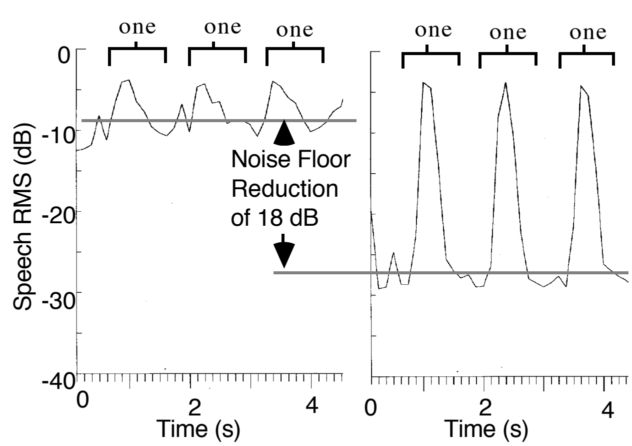


Fig. 4 — Measured frequency response from the prototype. The amplitude response approximates fairly well the theoretical response shown in Fig. 2. The near-field response is flat for frequencies below 2 kHz, and far-field response has a high-pass characteristic as in the theoretical response. The attenuation at one inch from the mouth location is as much as 24 dB which can be further raised by using a more refined summing amplifier and even better matched microphone cartridges.

4. NOISE REDUCTION PERFORMANCE

The microphone array presented in this paper improved the speech-to-noise ratio as much as 18 dB in the presence of E2C aircraft noise at an actual sound pressure level of 105 dB (Fig. 5).



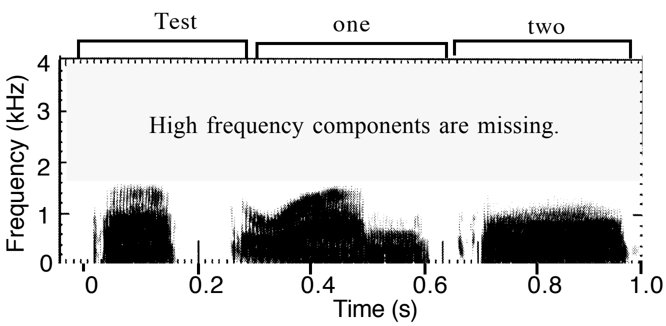
(a) No-noise cancellation b) Our microphone array

Fig. 5 - Histogram of speech RMS in the presence of E2C aircraft noise. As noted, our noise-canceling microphone improves the speech-noise ratio by 18 dB.

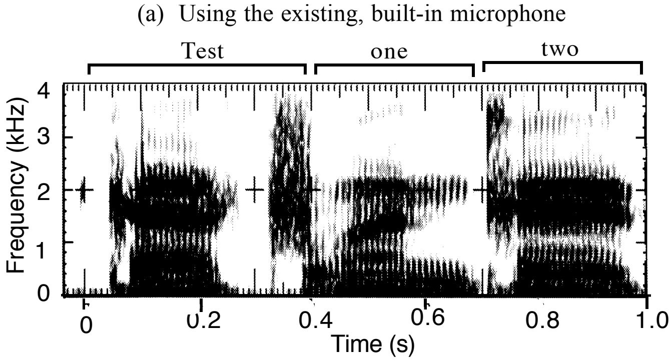
5. FACE MASK SPEECH IMPROVEMENT

Reverberations present in the face mask cause speech to sound muffled. Muffled speech lacks high-frequency components (Fig. 6a) which makes speech not too intelligible. This speech can be improved by recovering these high frequency components (Fig. 6b).

The reverberating speech within the mask can be viewed by the microphone array as more distant sound sources than speech emanating directly from the speaker's mouth. Since the microphone array provides greater attenuation for distant sound sources than nearby sound sources, the reverberations are largely suppressed and the speech becomes much more intelligible. Most remarkably, the array output contains the high frequencies which were missing in the reverberated speech (Fig. 6).



(a) Using the existing, built-in microphone



(b) Using the microphone array discussed in this paper

Fig. 6 - Speech spectra from the oxygen face mask. Note the presence of upper frequency components in Fig. 6(b) that are lacking in the muffled speech shown in Fig. 6(a). Both speech samples were spoken by the same person.

7. PREFERRED IMPLEMENTATION

There are two aspects of implementation which are beneficial to practical applications:

Automatic Noise Reduction Control

Note that the more powerful a noise-canceling microphone is, the more care that is needed to hold the microphone closely and steadily to the mouth. Otherwise, speech will fade in and out, diminishing the effectiveness of communication. The microphone array should be implemented in such a way as to automatically adjust the noise cancellation capability in accordance with the ambient noise level. In this way, the microphone will be effective in quiet, noisy, or intermittently noisy environments; such as the flight deck of an aircraft carrier.

Digital Implementation

We envision that this microphone array will be implemented digitally, with high-rate A/D converters at the front end. The digital implementation is preferred over the analog implementation because digital processors can detect small signal differences more accurately. Then, the noise reduction performance will be even better than what is shown in Fig. 2.

8. SUMMARY

An effective way of improving speech in noise, as well as within the face mask, is to reduce interference at the microphone. A desired characteristic of such a microphone is a large attenuation per source distance (20 dB or more at one inch away from the mouth).

For ambient noise reduction, our microphone array achieved an improvement of the speech-to-noise ratio that is close to 18 dB. When our microphone was used within the face mask, the high frequency contents of speech, which were lost because of reverberations, were almost completely restored. Potential improvements in the implementation of the design mentioned may lead to even better performance.

9. REFERENCES

1. W.A. Beaverson and A.M. Wiggins, "A Second-Order Gradient Noise Canceling Microphone Using a Single Diaphragm," *J. Acoust. Soc. Am.* **22**(5), 192-601 (1950).

2. A. Brouns, "Second-Order Gradient Noise-Canceling Microphone," *Proc. 1981 IEEE Int. Conf. Accoust. Speech Signal Process*., 786-789 (1981).

3. T.H. Crystal, "National Research Council Panel Study on Noise Reduction," *Proc. Mil. Speech Tech. '87*, 113-117 (1987).

4. J.S. Lim and A.V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. IEEE*, **67**, 1586-1604 (1979).

5. Olson, H, F., "Gradient Microphones," *J. Acoust. Soc. Am.* **17**(3), 192-198 (1946).

6. Olson, H.F., *Acoustical Engineering*, Professional Audio Journals, Inc., Philadelphia, PA, 1991.