Veterans Administration

Journal of Rehabilitation Research
and Development Vol. 24 No. 4
Pages 87–92

Effect of two-microphone noise reduction on speech recognition by normal-hearing listeners*

TERESA SCHWANDER and HARRY LEVITT**
Center for Research in Speech and Hearing Sciences,
Graduate School and University Center, City University of New York, New York, NY 10036

Abstract—An idealized 2-channel noise reducing adaptive filter of the type developed by Widrow requires that one channel contain noise only and that the microphones be fixed in position relative to the signal and noise sources. These conditions are unlikely to be met in a wearable hearing aid. In a typical situation, the microphones will be mounted in close proximity on a moving head in a room that is moderately reverberant. Experimental data have been obtained showing that, despite these deviations from the ideal conditions, significant improvements in speech intelligibility can be obtained using 2-channel adaptive filtering.

INTRODUCTION

The understanding of speech in a noisy environment is particularly difficult for hearing-impaired persons. As noted by Plomp (10), the speech-to-noise ratio required by a hearing-impaired person such that intelligibility is comparable to that for speech in quiet is significantly greater than the corresponding speech-to-noise ratio for a normal-hearing person. This effect is most pronounced for persons with sensorineural impairments. As a consequence, amplification of speech in noise provides little benefit to the sensorineurally hearing-impaired person since both the speech and noise are amplified leaving the speech-to-noise ratio unchanged. This is a particularly serious problem since the vast majority of hearing aid users have sensorineural impairments. Not surprisingly, one of the most common complaints about hearing aids is that these instruments are of little or no value in a noisy environment. The possibility of using modern signal processing techniques so as to reduce background noise is thus of great potential value, particularly if the signal processor can be made small enough to be incorporated in a wearable hearing aid.

Noise reducing or noise-stripping algorithms can be subdivided into two groups, those that are restricted to a single microphone input (single-channel systems) and those that have two or more inputs (multi-channel systems). A review of single-channel systems indicates that although modern signal-processing techniques can produce significant improvements in speech-to-noise ratio, concomitant improvements in speech intelligibility have not been obtained (7). An important recent development is the single-channel processor developed by Graupe et al. (5). This unit is small enough to fit into a conventional hearing aid and preliminary results obtained with this system have been favorable (11).

In contrast to single-channel systems, substantial improvements in speech-to-noise ratio and concomitant improvements in intelligibility have been obtained with multi-channel systems (1,3). A particularly promising approach is the two-channel adaptive

*The research reported here was supported by Grant No. GOO8302511 from the National Institute of Disability and Rehabilitation Research (NIDDR) to the Lexington Center, Jackson Heights, NY 11370.
**Reprint requests should be addressed to: Teresa J. Schwander, Veterans Administration Medical Center 561-126, Tremont Avenue, East Orange, NJ 07019.

filter system developed by Widrow et al. (16). The application of the Widrow approach to the hearing aid problem was first reported by Brey and Robinette (1), who obtained improvements in speech-to-noise ratio of at least 20 dB with correspondingly large improvements in intelligibility. There are, however, a number of practical constraints limiting the usefulness of two-channel adaptive filters with hearing aids and these need to be investigated.

The essential requirements of a two-channel adaptive filter for noise reduction are:

1) Two microphones must be used. One microphone, referred to as the *reference* microphone, picks up primarily background noise. The second microphone, referred to as the *primary* microphone, picks up both speech and noise.

2) An adaptive filter is required to modify the output of the reference microphone such that the difference between the primary input (speech plus noise) and the filtered reference input (mostly noise) is minimized. It can be shown that this difference signal consists of speech plus noise where the speech-to-noise ratio has been maximized. For further information on two-channel adaptive filters and their operation, see Widrow and Stearns (15), and the papers by Chabries et al. (4) and Weiss (14) in this issue.

There are at least three practical limitations to the use of two-channel adaptive filtering with hearing aids. The first is that, for a practical system, both microphones should be worn on the body, preferably on the head. This reduces the extent to which the reference microphone can be used to pick up the noise signal. That is, the reference input will contain both speech and noise. Any decrease in the noise-to-speech ratio at the reference input will reduce the speech-to-noise ratio at the output of the system. A second limitation is that room reverberation, or an increase in the number of noise sources, will reduce the effectiveness of the noise reduction system. The third limitation is that the adaptive filter needs time to adapt. This could cause problems if both microphones are mounted on the head and the head is moving relative to the speech and noise sources.

This paper is concerned with evaluating a two-channel adaptive filter for noise reduction subject to constraints typical of actual hearing aid use. The purpose of this present study was to evaluate the effect on speech recognition of a signal processor using two head-mounted microphones in a moderately reverberant room. The following questions were proposed:

1) Would a head-mounted directional reference microphone improve the noise-to-speech ratio sufficiently to improve speech recognition to the degree observed for an uncontaminated reference?

2) Would changes in microphone orientation, as with movement of the listener's head, reduce the improvement in speech recognition that could be obtained after processing?

PREPARATION OF TEST MATERIALS

Recordings of test stimuli were made in a room (18.5 ft. x 20 ft. x 9 ft.) with an average reverberation time equal to .41 seconds. This room was chosen to represent a typical amount of room reverberation, and is similar to the reverberant condition evaluated by Chabries et al. (3). Two recording microphones were mounted on the head of a listener who was seated in the center of the room. An omni-directional microphone (Knowles EA-1842) worn at the listener's right ear served as the primary microphone. A cardioid microphone (Beyer dynamic M201NC), mounted on top of the listener's head facing toward the rear, served as the reference microphone. This microphone array was selected to optimize the noise-to-speech ratio at the reference microphone with regard to the locations of speech and noise in the room (13). The large size of the cardioid microphone, chosen for its flat frequency response, necessitated mounting it atop the head rather than at the listener's ear.

Monosyllabic words (N.U. Auditory Test #6) and speech spectrum shaped noise were introduced into the room through two loudspeakers. Speech was presented from an azimuth of 0 degrees and noise from an azimuth of 180 degrees relative to the listener. The location of speech was selected to represent face-to-face communication, which is the ideal situation for an impaired listener to also use visual speech cues (10). The location of noise was chosen to maximize spatial separation of the speech and noise. The listener was seated in the center of the room at a distance of 8.5 feet from the two loudspeakers.

Speech was presented at an intensity of 72dB SPL measured one meter from the loudspeaker. This level represents an average intensity for a male talker (8). The intensity of the noise was that which resulted in a signal-to-noise ratio of 0 dB measured at the output of the primary microphone. This signal-to-noise ratio was chosen based on a preliminary study to result in approximately 50 percent word recognition by inexperienced normal-hearing listeners.

The output of each microphone was amplified, digitally processed by a pulse code modulator (Sony PCM-F1) and recorded on a two-track wideband recorder (Panasonic NV8420). This system provided high quality recordings that were limited only by the bandwidth and dynamic range of the microphone used.

Recordings were made for two conditions of head movement, no-head-movement, and moderate-head-movement. In the no-head-movement condition, the listener maintained her head position as stationary as possible. In the moderate-head-movement condition, the listener moved her head systematically from right to left by ±13 degrees and up and down by ±10 degrees.

Measurements of typical head movements were obtained prior to this study. A lightweight, narrow-beam flashlight was mounted over the right ear of a subject engaged in conversation. The test subject was seated 6.5 feet from a blank wall such that the movements of the light beam from the headworn flashlight were clearly visible on the wall. Excursions of the light beam were monitored and a record kept of the extreme excursions obtained over several minutes of lively conversation. These extreme excursions were found to correspond to angular movements of ±13 degrees in the horizontal direction and ±10 degrees in the vertical direction.

The reverberation time of the test room was measured for one octave bands of noise with center frequencies 250, 500, 1000, 2000, 4000, and 8000 Hz. Broadband noise bursts (2 seconds duration, 10 ms rise/fall time) were used as the test stimulus. These were generated by a Grason-Stadler white noise generator (GSC 901-B) the output of which was controlled by an electronic switch (GSC 829-C), and an interval timer (GSC 471-1). The noise bursts were amplified and played through a Wharfdale (W25) loudspeaker placed 4 feet from one wall. A sound level meter (B&K 2203) coupled to a standard one-octave-wide filter with adjustable center frequency (B&K 1613) was located 10.5 feet from the signal source. This distance was derived from the critical distance formula of Peutz (9). Level recordings showing the rate of decay of each noise burst were obtained using a graphic level recorder (B&K 2305). Reverberation time, defined as the time taken for the signal to decrease 60 dB from its original intensity, was calculated using a special protractor (B&K SC 2361). Reverberation times were obtained for the one-octave filter set to center frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz. The measured reverberation times were 0.35, 0.30, 0.36, 0.50, 0.48, and 0.45 seconds, respectively. The average reverberation time for the room was thus 0.41 seconds.

The two-channel recordings obtained from the primary and reference microphones were played back through a PCM decoder (Sony PCM-F1) into a two-channel adaptive filter (Adaptive Digital Systems Modular Adaptive Signal Processor) programmed so as to implement the algorithm developed by Widrow et al. (15). A filter length of 800 taps at a sampling rate of 10,000 Hz was used.

The choice of a 10 kHz sampling rate required that the audio signals processed by the system be limited to a bandwidth of just under 5 kHz. This bandwidth is comparable to that typically used in conventional hearing aids.

The choice of an 800-tap filter represented a compromise between 1) a long filter with good noise-reducing properties and a slow rate of adaptation, and 2) a short filter with poor noise-reducing properties and a rapid rate of adaptation. Data in the companion paper by Weiss (14) show that for the test room considered in this study, an 800-tap filter provides close to the maximum noise reduction within the time taken for the head to move from one extreme position to another (about 1 second) in lively conversation. The subjective judgements of two experienced listeners also supported the choice of an 800-tap filter as providing the best reduction in background noise for the experimental conditions considered in this study.

A set of test recordings was made for each of the four experimental conditions:

1) no-head-movement, unprocessed;

2) moderate-head-movement, unprocessed;

3) no-head-movement, processed to reduce noise; and

4) moderate-head-movement, processed to reduce noise.

Journal of Rehabilitation Research and Development Vol. 24 No. 4 Fall 1987

Table 1.
Percent Word Recognition

Subject	Unprocessed: No Head Movement	Unprocessed: Moderate Head Movement	Processed: No Head Movement	Processed: Moderate Head Movement
[illegible]L	18	26	68	56
LH	26	36	74	52
LW	32	20	64	56
AB	40	42	74	64
MB	54	42	88	84
Mean	34.0	33.2	73.6	62.4
Std Error	6.2	4.4	4.1	5.7

These recordings were played to five normal-hearing listeners, ages 29 to 49, who served as subjects. Stimuli were presented monaurally using a standard TDH-39 headphone. Word lists and listening conditions were randomized across subjects. Subjects were required to write down their responses.

RESULTS

Word recognition scores for the five subjects on each of the four experimental conditions are shown in **Table 1**. A repeated measures analysis of variance was performed, the results of which are shown in **Table 2**. Since the raw data were in the form of percentages, an inverse sine transformation was used to stabilize the error variance (2).

The results of the analysis showed that processing to reduce noise produced a significant improvement in word-recognition scores, from 33.6 to 68.0 percent, on the average. Head movement had a smaller, but statistically significant effect. These data are summarized in **Figure 1**. Note that there is no significant difference between the two unprocessed conditions but, for the processed signals, the score for the moderate-head-movement condition is significantly ($p<0.01$) below that for no-head-movement form. Tests for statistical significance were performed using the inverse sine transformation.

The analysis also showed large inter-subject differences. These ranged from an average score of 67 percent for the best subject to 42 percent for the poorest subject. No sigificant interactions were observed between subjects and the various test conditions.

Measurements of the reduction in noise level resulting from the use of the two-channel adaptive filter were also obtained for the no-head-movement condition. The reduction in noise level for an 800-tap filter at a sampling rate of 10,000 Hz was 7.0 dB. For the condition involving moderate-head-movement the corresponding reduction in noise level was 4.5 dB. These data represent processing effects on noisy speech in the moderately reverberant room used in this study.

Table 3 provides some comparable data obtained by Weiss (13) showing how the amount of noise reduction is affected by filter length and room reverberation for the fixed-head condition. These measurements were obtained using the microphones mounted on an anthropometric manikin (KEMAR). Data are shown for the moderately reverberent room used in this study as well as a sound-treated test room with a low reverberation time (0.2 seconds).

Table 2.
Results of Analysis of Variance. A repeated-measures model has been used. The observed proportions were subjected to the inverse sine transformation $y = 2 \sin^{-1}\sqrt{p}$ prior to the analysis.

Source of Variation	Degrees of Freedom	Mean Squares	F-Ratio	Significance Level
Movement (M)	1	0.081	14.2	0.021
Processing (P)	1	2.637	190.2	0.001
M × P	1	0.065	2.1	0.218
Subject(s)	4	0.214	7.0	0.045
M × S	4	0.006	0.2	0.933
P × S	4	0.014	0.4	0.769
M × P × S	4	0.031	–	

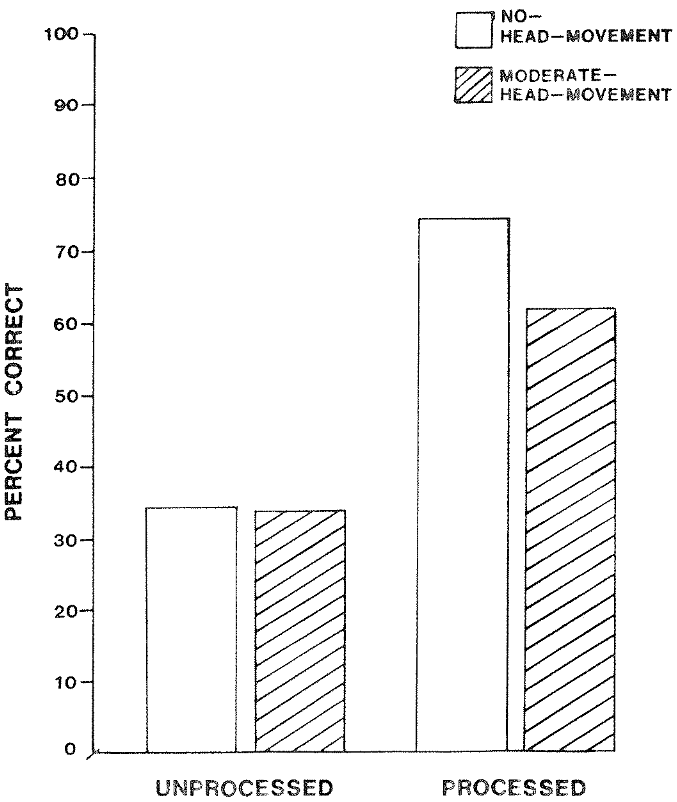

Figure 1.
Average word recognition performance by five normal hearing listeners.

DISCUSSION

Two important conclusions can be drawn from the results of this study. The first is that two-channel adaptive filtering based on the Widrow algorithm can provide a significant improvement in speech intelligibility, even under conditions that differ substantially from those for which the Widrow technique was developed. Under ideal conditions, the reference input should consist of noise only. This is highly unlikely to occur in a practical hearing aid in which both microphones are mounted on the body, preferably on the head. Although the use of two head-mounted microphones resulted in the reference input containing both speech and noise, the difference in speech-to-noise ratio between the reference and primary inputs was still sufficiently large to produce a significant reduction in the background noise.

A second major problem in applying the Widrow technique to everyday hearing aid use is that the acoustic environment is likely to include some reverberation, as would occur in a typical room. The two-channel adaptive filter does not work well in a reverberent environment because the acoustic reflections serve to reduce the differences between the reference and primary inputs; i.e., the speech-to-noise ratio at the primary input is reduced, as is the noise-to-speech ratio at the reference input.

Table 3.
Noise attenuation as a function of filter length and adaptation time for a reverberant room (RT = 0.41 sec) and a sound treated room. Sampling rate was equal to 12,000 Hz.

	Adaptation time	Filter Length (# of taps)				
		100	200	400	600	1000
Sound treated room						
18 dB 18 dB	1 sec	16 dB	17.5 dB	18 dB	18 dB	18 dB
32 dB 32 dB	14 sec	21 dB	28 dB	32 dB	32 dB	32 dB
Reverberant room						
7 dB 7.5 dB	1 sec	4 dB	5 dB	6 dB	7 dB	7.5 dB
9 dB 13 dB	14 sec	5 dB	6 dB	8 dB	9 dB	13 dB

The results of the present investigation show that, despite these substantial deviations from the idealized condition of noise only at the reference input, it is still possible to obtain a significant improvement in speech-to-noise ratio and a concomitant increase in intelligibility. The improvement in speech-to-noise ratio for the no-head-movement condition was found to be roughly 7.0 dB. For the monosyllabic word test used in the study, this improvement resulted in an average increase of 40 percentage points in word recognition score. For speech materials having a steeper performance intensity function (e.g., sentences), the gain in percent intelligibility would have been even greater.

The second important conclusion is that head movement does reduce the effectiveness of the noise reduction process, but by a relatively small amount. Although the amount of head movement was fairly large (extreme excursions of the head having been used in making the recordings), the average reduction in word recognition score was only 10 percentage points. It should also be noted that head movements took place while recording the test stimuli, and not while listening to the processed speech signals. The latter condition is closer to real-life listening and contains additional cues that could be helpful to the listener in paying attention to the speech signal. Although the conditions considered in this study deviate significantly from the idealized conditions for a 2-channel adaptive filter, the results nevertheless indicate that signal processing of this type can still be effective when the requirements of an uncontaminated noise channel and fixed microphone positions are not met.

It should be remembered that an 800-tap filter was chosen as the best compromise between good noise reduction with a slow rate of adaptation and poor noise reduction with a rapid rate of adaptation. This choice of filter length was made for the specific experimental conditions considered in this study. It may be that, for rooms with different reverberation times or for different rates of head movement, another filter length may be optimum.

An important practical advantage of the two-channel adaptive filtering for noise reduction is that extensive processing of the speech signal is not required. In the original Widrow procedure, for example, the reference input containing noise only is processed and then subtracted from the primary input containing speech plus noise. As a consequence, the speech signal is not distorted by imperfections in the signal processor. A serious problem in many noise reduction systems is that, although the signal processing technique may improve speech-to-noise ratio, there also are concomitant distortions of the speech signal such that the expected gain in intelligibility resulting from the improved speech-to-noise ratio are lost because of these distortions. This problem is especially common in single-channel noise reduction systems. [See, for example, Lim and Oppenheim (7) and Levitt et al. (6)].

It was not the purpose of this study to investigate whether the reduction in background noise of the processed signals produced the same gain in percent word recognition as simply attenuating the noise without any processing. The measured improvements in speech-to-noise ratio and gain in percent word recognition obtained in this study are consistent with measurements obtained in other experiments on the slope of the performance intensity function for these test materials (12). That is, the improved speech-to-noise ratio produced by the two-channel adaptive filter has resulted in an increase that is essentially the same as that to be expected from simply increasing the signal-to-noise ratio without any signal processing. Thus, there does not seem to be a reduction in percent word recognition resulting from the signal processing operation *per se*.

REFERENCES

1. **Brey RH and Robinette MS:** Noise reduction in speech using adaptive filtering II: Speech intelligibility improvement with normal-hearing and hearing-impaired students. Paper presented at 103rd meeting of ASA, Chicago, Illinois, 1982.
2. **Brownlee KA:** *Statistical Theory and Methodology in Science and Engineering*, Second Edition. New York: John Wiley and Sons, 1965.
3. **Chabries D, Christiansen R, Brey R and Robinette M:** Application of the LMS adaptive filter to improve speech communication in the presence of noise. Paper presented at IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, France, 1982.
4. **Chabries DM, Christiansen RW, Brey RH, Robinette MS, and Harris RW:** Application of adaptive digital signal processing to speech enhancement for the hearing impaired. *J Rehabil Res Dev* 24(4) (this issue), 1987.
5. **Graupe D, Grosspietsch JK, and Basseas SP:** A single-microphone-based self-adaptive filter of noise from speech and its performance evaluation. *J Rehabil Res Dev* 24(4) (this issue), 1987.
6. **Levitt H, Neuman A, Mills R, and Schwander TM:** A digital master hearing aid. *J Rehabil Res Dev* 23(1):79-87, 1986.
7. **Lim J and Oppenheim A:** Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE* 67(12):1586-1604, 1979.
8. **Pearsons K, Bennett R, and Fidell S:** Speech levels in various environments. *BBN Report,* No. 3281, October, 1976.
9. **Peutz UMA:** Articulation loss of consonants as a criteria for speech transmission in a room. *J Audio Eng Soc* 19:915-919, 1971.
10. **Plomp R:** Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am* 63(2):533-549, 1978.
11. **Stein L and Dempesy-Hart D:** Listener-assessed intelligibility of a hearing aid self-adaptive noise filter. *Ear Hear* 5(4):199-204, 1984.
12. **Studebaker GA:** Personal communication, 1987.
13. **Weiss M:** Personal communication, 1987.
14. **Weiss M:** Use of an adaptive noise canceler as an input preprocessor for a hearing aid. *J Rehabil Res Dev* 24(4) (this issue), 1987.
15. **Widrow B and Stearns S:** *Adaptive Signal Processing*. Englewood Cliffs, N.J.: Prentice-Hall, 1985.
16. **Widrow B, Glover J, McCool J, Kaunitz J, Williams, C, Hearn R, Zeidler J, Dong E, and Goodlin, R:** Adaptive noise cancelling: Principles and applications. *IEEE Proceedings* 63:1692-1716, 1975.