Veterans Administration

Journal of Rehabilitation Research
and Development Vol. 24 No. 4
Pages 271–276

# The short-time articulation index

JAMES M. KATES*
*SIGNATRON, Inc., Lexington, Massachusetts 02173*

**Abstract**—In this paper we introduce the concept of the short-time articulation index. This is a procedure for calculating a time-varying articulation index from data on a block-by-block basis. The short-time articulation index can be used to give a running measure of the speech intelligibility for an adaptive noise-cancellation system as it converges. We present an algorithm for calculating the short-time articulation index and give some examples of its use.

## INTRODUCTION

The conventional method for assessing the performance of a noise-cancellation system is to compute the noise reduction in dB. For speech, one problem with this method is that it measures only the change in the total noise power and ignores the effects of the noise spectrum on speech intelligibility. Speech intelligibility is best measured by conducting tests with human listeners, but procedures such as the articulation index (3, 4) have been developed that give repeatable results that can be related to intelligibility test scores (5). Although only a generalized approximation, the articulation index (AI) will give a better indication of the effect of background noise on speech intelligibility than simply considering the speech-to-noise ratio.

The classical articulation index calculation gives a single number based on measurements of the long-term noise power spectrum. This is useful for evaluating a stationary system, but inadequate for an adaptive system that may be changing with time. We have developed a simplified short-time articulation index calculation that is suitable for evaluating adaptive noise-cancellation systems. The procedure is based on the 20-band AI calculation procedure of French and Steinberg (3), and incorporates upward spread of masking as described by Kryter (4). This calculation is an approximation to the classical AI procedure because shaped Gaussian noise rather than speech is used as the system input, remote masking is not included, and the calculation is performed short-time on a block-by-block basis. That is, the signal to be analyzed is subdivided into a sequence of time windows and a separate AI analysis is performed for each time window. The size of each block (time window) is a parameter under the control of the experimenter.

## ADAPTIVE PROCESSING

A block diagram of a generic two-sensor adaptive noise-cancellation system, such as the one described by Widrow et al. (8), is shown in **Figure 1**. The primary signal p(n) consists of speech corrupted by additive noise, giving $p(n) = s(n) + q(n)$, where s(n) is the speech and q(n) the noise. The noise reference

* Current affiliation: Siemens Hearing Instruments, Inc., Piscataway, NJ 08854.

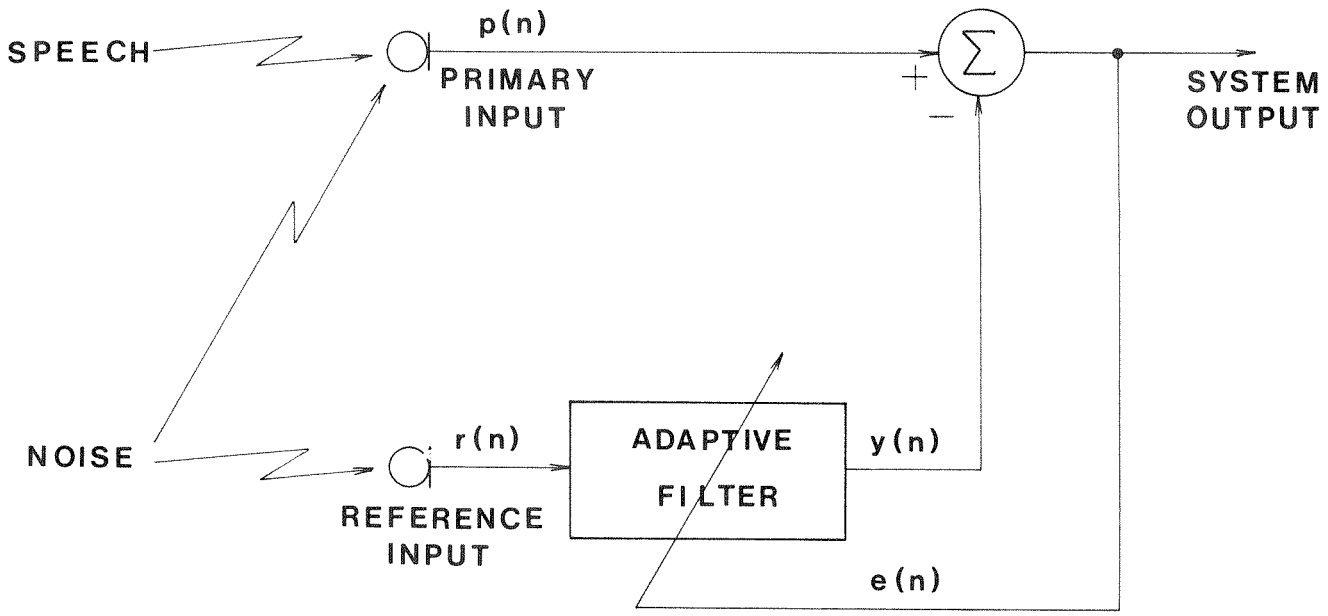


**Figure 1.**
Adaptive noise-cancellation system.

r(n) is obtained from a separate sensor and is filtered to give y(n). The adaptive filter weights are updated to make y(n) as close a match as possible to the noise q(n) that is corrupting the primary signal. Subtracting the filtered noise reference from the primary signal results in the system output

$$e(n) = p(n) - y(n)$$

which is also the error signal used to govern the adaptation of the filter weights.

The conventional performance criterion for an adaptive system is the signal-to-noise ratio computed on a block-by-block basis. Assuming that the processing is being simulated on a computer, we have access to speech signal s(n), the noise q(n) corrupting the speech signal, the noise reference r(n), and the processed output e(n). We can therefore subtract the input speech from the processed output to give the residual noise component after processing, and thus compare the noise power after the adaptive cancellation with the noise power corrupting the primary signal. The ratio of those two noise levels gives the amount of noise reduction, which is expressed in dB. The noise-reduction calculation is typically performed on a block-by-block basis in order to track the convergence of the adaptive algorithm.

## SHORT-TIME AI CALCULATION

The AI calculation is based on the ratio of the speech peaks to the average noise power level in each of 20 frequency bands that contribute equally to speech intelligibility. The original procedure assumed long-term averaging of both the speech and noise statistics. We have modified the basic procedure to permit analysis on a short-time block-by-block basis.

An actual speech signal is nonstationary and can vary considerbly from one analysis block to the next, so our first modification has been to replace the speech with a stationary noise signal having a spectrum similar to that of speech. This "pseudo-speech" signal is white Gaussian noise that has been passed through a one-pole lowpass filter having a cutoff frequency of 400 Hz. The speech peak levels required for the AI calculation are then approximated by adding 12 dB to the rms pseudo-speech level in each AI analysis band; the 12-dB peak-to-average ratio is indicated by Kryter (4).

The second change that we have made is in the AI analysis bands. This was done to accommodate the system sampling rate and to permit the use of the FFT algorithm in computing the signal and noise spectra. The AI analysis bands are therefore ap-

**Table 1.**
AI bands for a 64-point FFT, 10-kHz sampling rate.

| FFT Bin Index | Center Frenquency, Hz | AI Band |
|---|---|---|
| 0 | 0 | |
| 1 | 156 | |
| 2 | 312 | 1 |
| 3 | 469 | 2 |
| 4 | 625 | 3 |
| 5 | 781 | 4 |
| 6 | 937 | 5 |
| 7 | 1094 | 6 |
| 8 | 1250 | 7 |
| 9 | 1406 | 8 |
| 10 | 1562 | 9 |
| 11 | 1719 | 10 |
| 12 | 1875 | 11 |
| 13 | 2031 | 12 |
| 14 | 2187 | |
| 15 | 2344 | 13 |
| 16 | 2500 | |
| 17 | 2656 | 14 |
| 18 | 2812 | |
| 19 | 2969 | 15 |
| 20 | 3125 | |
| 21 | 3281 | 16 |
| 22 | 3437 | |
| 23 | 3594 | |
| 24 | 3750 | 17 |
| 25 | 3906 | |
| 26 | 4602 | |
| 27 | 4219 | |
| 28 | 4375 | 18 |
| 29 | 4531 | |
| 30 | 4687 | |
| 31 | 4844 | |
| 32 | 5000 | |

proximated by one or more discrete points (bins) in the discrete Fourier transform (FFT). The actual division of the spectra into analysis bands thus depends on the system sampling rate and the length of the data segments used for the short-time AI calculations. An example of this division is given in **Table 1** for a sampling rate of 10 kHz and data segments of 64 points. Here we have only 18 analysis bands, so the AI calculation must be modified in order for the ideal system performance to give an AI of one.

The short-time AI calculation thus proceeds in the following steps.

1. Excite the system using the lowpass-filtered noise to represent the speech signal and use any desired signal for the corrupting noise.
2. Divide the processed output into blocks, and for each block separate the signal into the pseudospeech and the residual noise.
3. Perform FFT's on the pseudospeech and residual noise sequences and group each spectrum into the AI bands as shown in **Table 1**.
4. Add 12 dB to the rms pseudospeech levels to approximate the speech peaks.
5. Modify the noise spectrum to include spread of masking. The magnitude of the spread-of-masking effect was obtained from Kryter (4). We have used only upward spread of masking in our calculations.
6. Compute the speech peak-to-noise ratio in dB in each AI analysis band. Values above 30 dB are set to 30, and values below 0 dB are set to 0.
7. Sum the ratios and divide by 30 times the number of analysis bands to give an AI score between 0 and 1.

## EXAMPLE

We have used the short-time articulation index to investigate the performance of an adaptive noise-cancellation system. The processing was simulated on a general-purpose computer, giving us acess to all of the necessary signals for performing the short-time AI calculations. The short-time AI and SNR were computed to give a comparison between these two performance measures.

The adaptive processing was tested with simulated noise and reverberation. The noise reference signal consisted of white Gaussian noise passed through a one-pole lowpass filter having a cutoff at 400 Hz. This spectral shape was chosen to give a long-term spectral shape similar to that of speech babble and many mechanical noises. The primary signal was the pseudospeech described in the previous section plus uncorrelated intereference consisting of the noise reference signal convolved with the early reflections from a small room. The reverberation was modeled as the direct sound plus the first 9 reflections from the walls of a room 2 meters wide by 3 meters long, giving an impulse response of 114 samples at the 10-kHz sampling rate. The pseudospeech was combined with the reverberated noise to give a 0 dB signal-to-noise ratio.

The adaptive processing was implemented using the frequency-domain approach of Ferrara (2); discussions of related frequency-domain adaptive proc-

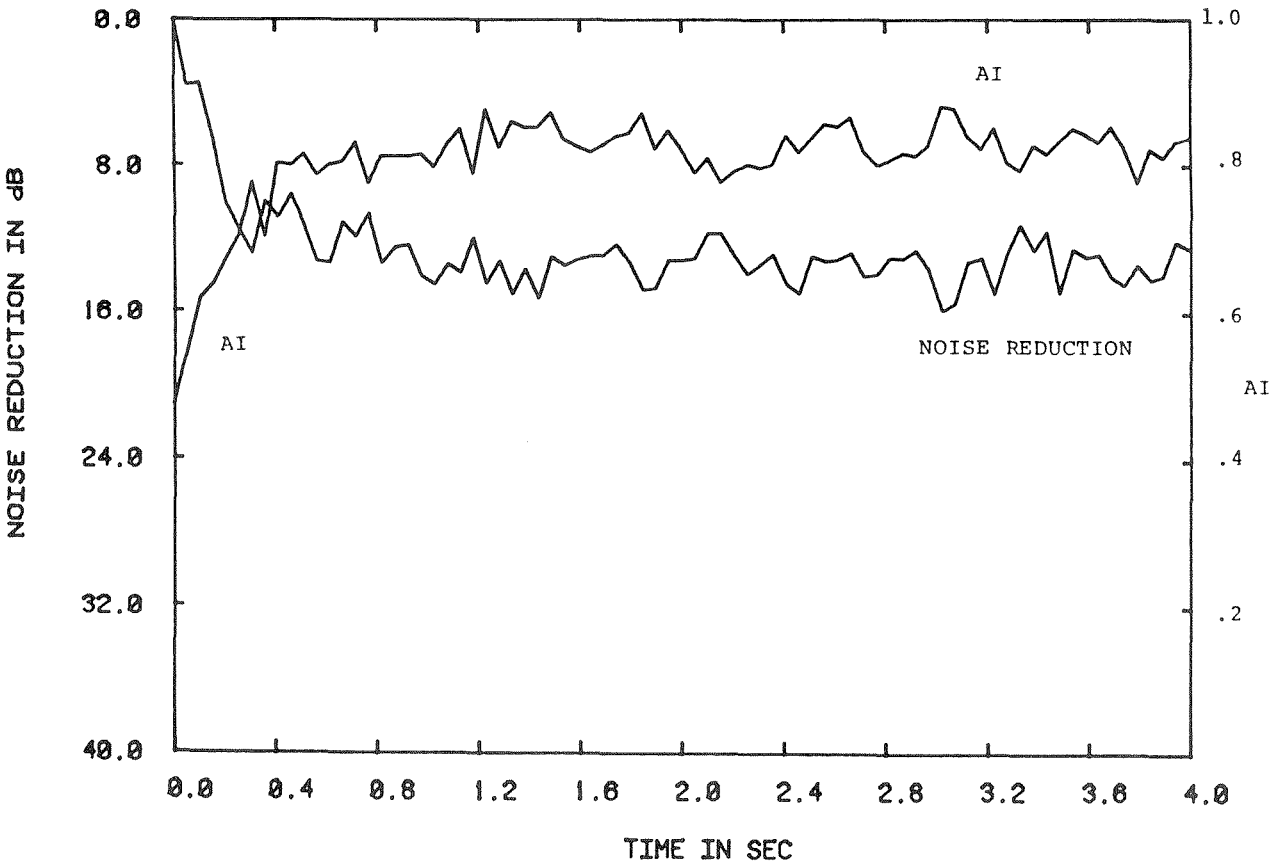


**Figure 2.**
Adaptation behavior: short time constant, fixed microphones.

essing can also be found in Mansour and Gray (6) and Clark, Parker, and Mitra (1). The algorithm used for updating the adaptive filter weights was the fast Kalman algorithm of Ogue, Saito, and Hoshiko (7). The adaptive filter length for the processing was set to 128 samples in order to accommodate the impulse-response duration of the simulated reverberation.

We first simulated the situation where the primary signal and noise reference microphones are fixed in position. **Figures 2** and **3** show the results of the processing for two different values of the adaptation time constant. In **Figure 2** we have a short time constant, and the processing shows rapid convergence but relatively poor asymptotic performance. The short-time AI curve is very nearly the mirror image of the noise-reduction curve, which is typical behavior for moderate amounts of noise reduction.

In **Figure 3** we have a long time constant, and the processing shows slow convegence but very good asymptotic performance. The short-time AI is again an approximate mirror image of the noise-reduction, but here the AI curve is smoother. This is especially true as we approach the asymptote at times beyond 2.0 seconds, since as the AI score approaches 1.0 it is only mildly affected by the small fluctuations in noise level that are shown by the noise-reduction curve.

We then simulated a situation where the primary microphone is allowed to move but where the reference microphone is still fixed. This corresponds, for example, to a head-mounted speech microphone as is commonly used in voice communications systems, but where the noise reference microphone is permanently mounted. The head motion was modeled as a sinusoidal oscillation of the primary microphone position along the long axis of the room, with the reflection time delays adjusted to track the changing microphone location. Thus the assumed sinusoidal motion was converted into a modulation of the tap delays used to simulate the reverberation. The head motion in our example was set to 0.1 m peak-to-peak displacement at a frequency of 1.0 Hz.

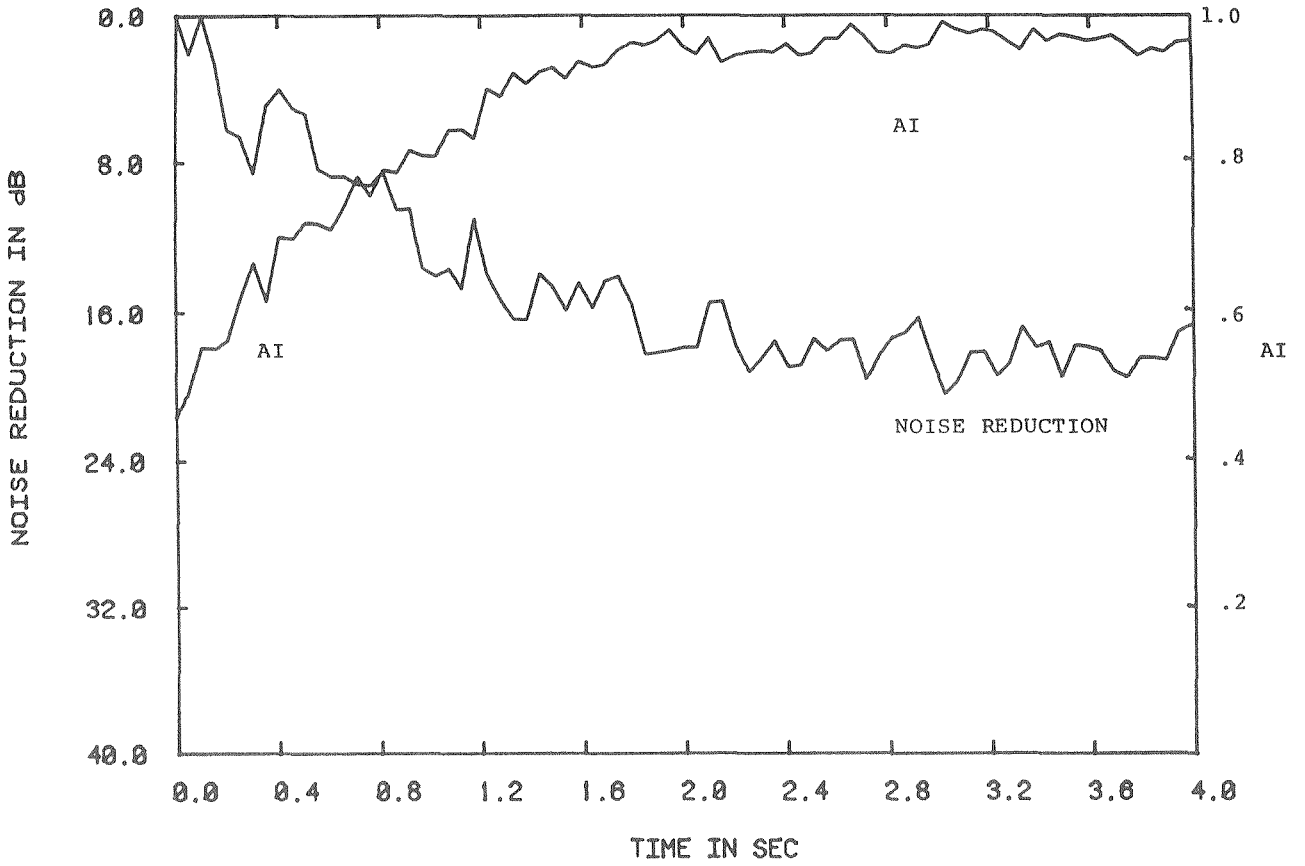


**Figure 3.**
Adaptation behavior: long time constant, fixed microphones.

The results for the simulated head motion are presented in **Figure 4** for the same adaptation time constant as was used for the curves in **Figure 3**. As we can see, the simulated head motion severely degrades the system performance. The asymptotic noise reduction has gone from about 18 dB to about 9 dB, and the asymptotic short-time AI has gone from about 0.97 to less than 0.7. And even though the noise-reduction curve shows an improvement in SNR for the adaptive system, the AI scores give a strong indication that very little improvement would actually be realized for voice communications. This suggests that most of the noise reduction is at frequencies that are not important for speech intelligibility, or that upward spread of masking has reduced the net advantage of the adaptive noise cancellation.

## CONCLUSIONS

We have presented the short-time articulation index, a new technique for evaluating the performance of adaptive noise-cancellation systems. The short-time AI takes into account the effects of the noise spectrum as well as the noise intensity in determining the quality of a time-varying voice communications system. The calculation procedure is based on the classical articulation index procedure, with modifications to permit short-time computation using the FFT. The only constraint in using the short-time AI in testing a system is that the lowpass filtered pseudospeech signal should be used as the input to give a stationary signal having a spectrum similar to that of speech.

The short-time articulation index, when used in conjunction with the output SNR, will give a much more complete picture of the performance of an adaptive system. This is especially true where small amounts of noise reduction are being obtained, since the short-time AI will directly indicate if the noise reduction is adequate to improve voice communications. The short-time AI will be less useful when large amounts of noise reduction are possible, since it will approach its asymptote of 1.0 and not reflect

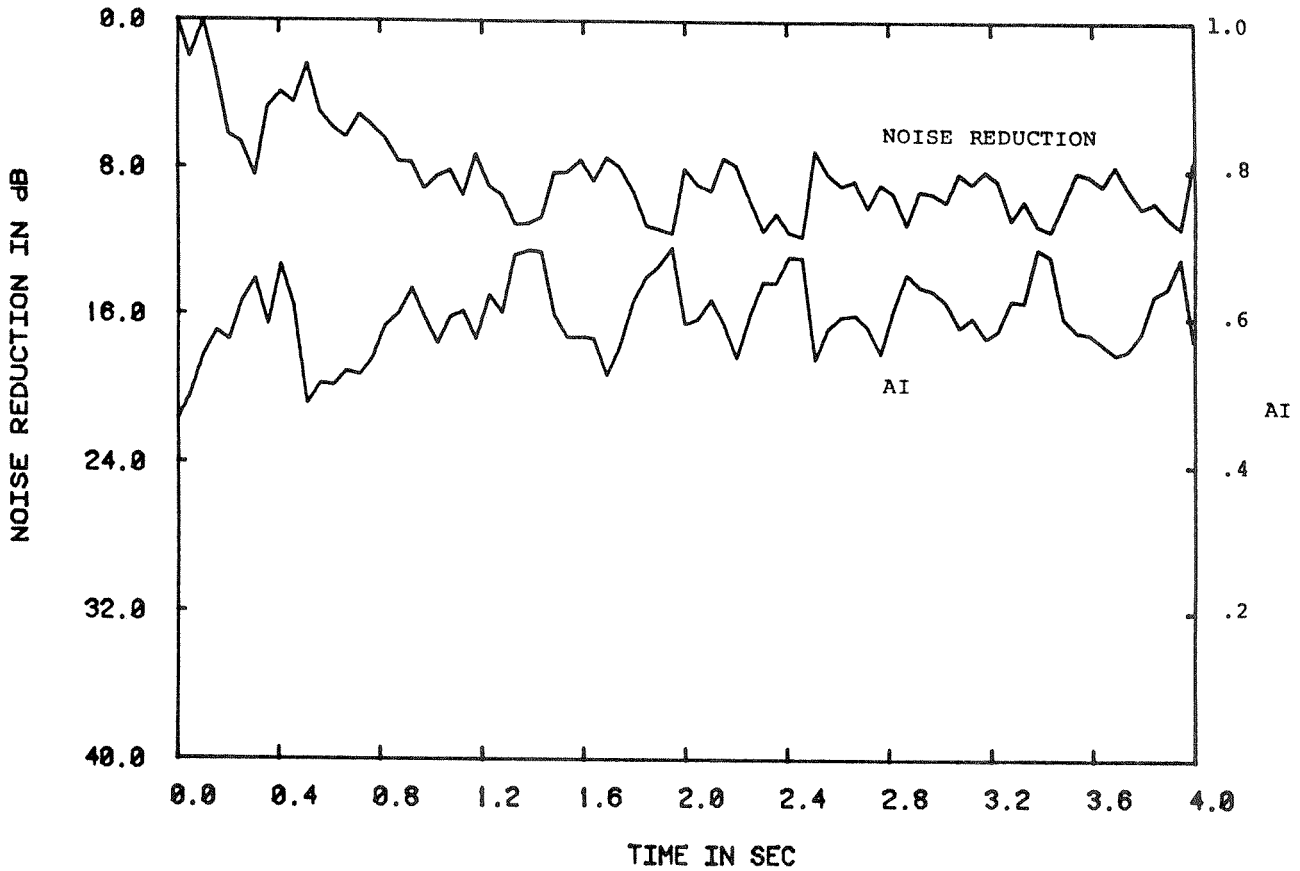


**Figure 4.**
Adaptation behavior: long time constant, oscillating primary microphone.

the additional small improvements in SNR that may result from the processing. In those cases, however, the short-time AI will indicate when further processing is not needed because of the quality of the voice communications that have already been achieved. Thus the short-time AI scores can also serve as a criterion for determining the degree of processing necessary to reach a desired level of speech intelligibility.

## REFERENCES

1. **Clark GA, Parker R, and Mitra SK:** A unified approach to time- and frequency-domain realization of FIR adaptive digital filters. *IEEE Trans,* ASSP-31(5): 1073-1083, 1983.
2. **Ferrara ER:** Fast implementation of LMS adaptive filters. *IEEE Trans,* ASSP-28(4):474-475, 1980.
3. **French NR and Steinberg JC:** Factors governing the intelligibility of speech sounds. *J Acoust Soc Am* 19:90-119, 1949.
4. **Kryter KD:** Methods for the calculation and use of the Articulation Index. *J Acoust Soc Am* 34(11):1689-1697, 1962.
5. **Kryter KD:** Validation of the Articultion Index. *J Acoust Soc Am* 34(11):1698-1702, 1962.
6. **Mansour D and Gray AH, Jr:** Unconstrained frequency-domain adaptive filter. *IEEE Trans,* ASSP-30(5):726-734, 1982.
7. **Ogue J, Saito, CT, and Hoshiko Y:** A fast convergence frequency-domain adaptive filter. *IEE Trans,* ASSP-31(5):1312-1314, 1983.
8. **Widrow BJ, Glover R, Jr, McCool JM, Kaunitz J, Williams CS, Hearn RH, Zeidler JR, Dong E, Jr and Goodlin RC:** Adaptive noise cancellation: Principles and applications. *Proc IEEE,* 63(12):1692-1716, 1975.