Noise reduction via signal processing: (1) Strategies used in other industries

By Donald J. Schum

A principal goal in hearing aid development for many years has been to minimize the negative effects of background noise on the wearer. Microphone approaches to noise control (directional and FM) improve the user's performance in background noise under the appropriate conditions. Sometimes, however, for various technical and practical reasons, microphone approaches are not feasible or they are unable to alleviate communication difficulties sufficiently. In such cases, there is a demand for advanced signal processing circuitry that can remove the noise component of a mixed signal (usually speech-plus-noise) input.

This two-part article will discuss first how the problem of hearing in noise has been addressed outside the hearing aid industry and then, next month, review the approaches our industry has taken. To demonstrate many of the noise-reduction techniques described in this article, we have compiled a Noise Reduction Demonstration CD with 10 sound samples. The samples provide an audio accompaniment to the descriptions and figures in this article. The CD is available free of charge from Oticon, Inc., by calling 800/526-3921.

The hearing aid industry is not alone in wishing to use signal processing to improve performance in noise. Other, far larger interests, including the military, intelligence agencies, and the huge telecommunications industry, share this objective.

A simple web search reveals many private companies that specialize in noise-reduction signal processing approaches. In general, the commercially available noise-reduction approaches fall into two major categories: *spectral subtraction* and *phase cancellation*. Almost all these applications are implemented digitally, since digital signal processing (DSP) is far more flexible than analog processing.

SPECTRAL SUBTRACTION

The basic approach of spectral subtraction is to identify the frequency characteristics of the competing signal and subtract it from the broadband spectrum of the speech signal. This requires as accurate as possible an estimate of the spectral composition of the noise. When those characteristics are known, it is relatively simple to use inverse filtering to remove only those frequencies from the signal.

It is obvious, however, that *the effectiveness of this approach is directly related to the width of the band of the competing signal.* Remember, this filtering will remove every signal component in the targeted frequency region, both the unwanted noise and the speech energy. Speech information is spread across frequencies ranging from below 200 Hz to above 5000 Hz, with most of the important information concentrated between approximately 1000 and 3000 Hz. If the competing noise has a narrow bandwidth (let's say, less than an octave), then filtering this part of the spectrum will leave sufficient speech information for complete recognition of the message. However, as the bandwidth of the competition increases,

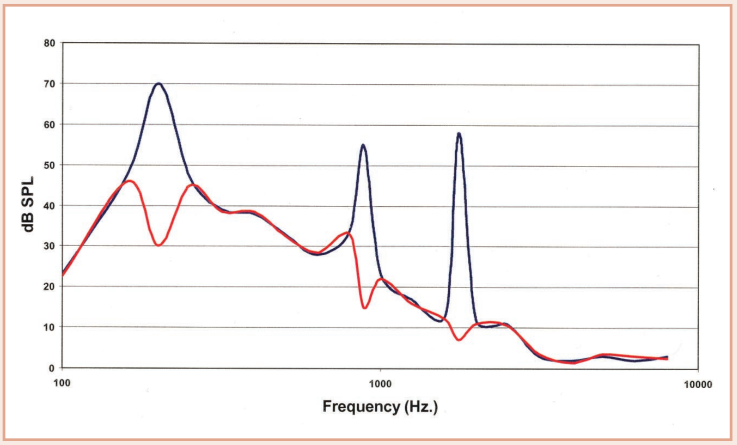

Figure 1. *The long-term spectra of a speech signal mixed with a three-component chord before (blue) and after (red) filtering.*

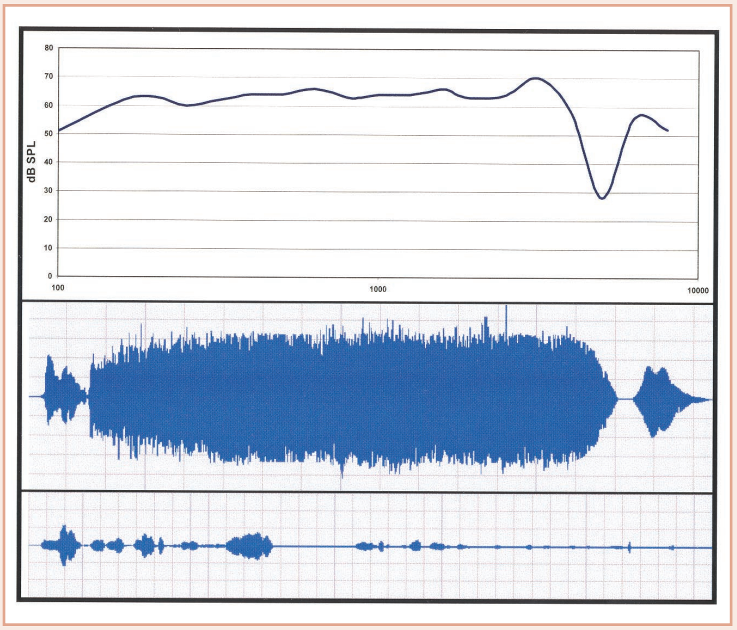

Figure 2. *The long-term spectra of the power drill used in auditory sample 2 (upper panel), the waveform before filtering (middle panel), and after filtering (lower panel).*

the noise filtering removes an increasing portion of the critical speech information.

Audio sample 1 on the CD is designed to show the effects of inverse filtering (spectral subtraction) on a narrow bandwidth noise source. In this example, speech is mixed with a three-part chord (like a sustained chord on an organ), consisting of a fundamental of 220 Hz and harmonics at 880 and 1760 Hz (see Figure 1, blue line). The frequency components of the chord were identified and an inverse filter with notches at those three frequencies was applied. The result (Figure 1, red line) is almost a complete elimination of the noise without any observable effect on speech understanding.

Audio sample 2 is another example of inverse filtering. However, in this case, the noise source was a drill that had a broad bandwidth spectrum (from 150 through 10,000 Hz, with a null around 5500 Hz—see Figure 2, upper panel). The filter necessary to eliminate this nose is so broadband that it includes nearly the entire speech spectrum. The middle panel of Figure 2 shows the waveform of the speech and drill before filtering, while the bottom panel shows the effect of the filter. The filter eliminated the noise, but it also eliminated nearly the entire speech signal. Together, Figures 1 and 2 clearly demonstrate that the usefulness of inverse filtering is directly related to the bandwidth of the competition.

In other industries, spectral subtraction is used widely to remove stable, nonspeech signals. For example, telecommunication companies use filtering to remove hiss or hum from transmission channels. The military has used inverse filtering to eliminate the predictable noise within a vehicle (e.g., a jet fighter or tank). Other industries have typically *not* applied this approach to the removal of unwanted speech or other broadband competition because of the problem that we discussed above related to the breadth of the spectrum. However, despite this known limitation, spectral subtraction in various forms has been the primary approach tried by the hearing aid industry over the past two or three decades. These past attempts will be discussed in part two of this article, next month.

PHASE CANCELLATION

A more sophisticated approach to noise removal makes use of the mathematically proven fact that if a waveform is copied, reversed 180 degrees in phase, and then added back into the original waveform, the sound will be completely canceled. Figure 3 demonstrates this effect with a simple waveform.

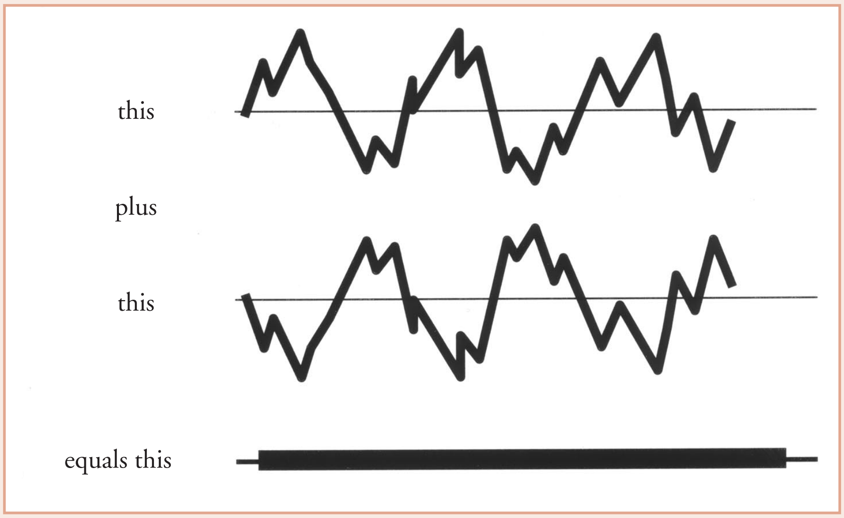

Figure 3. *This example shows how a phase-reversed waveform, once added back to the original, completely cancels the original sound.*

Phase cancellation is used in directional microphones (both traditional, two-port acoustic applications and more recent, twin-microphone electronic versions). The signal entering the rear port is delayed, reversed, and then added to the signal reaching the front port. The portion of the signal arising from the rear is thus canceled.

This approach has also been used in so-called "noise cancellation" headphones, which have recently become commercially available.

Phase cancellation is a very powerful signal processing approach, as it can separate one voice from various types of broadband competition—even speech. However, to be effective, a crucial requirement must be met: *The system needs access to the precise waveform (not just the spectral content) of the competition that is completely independent of the signal of interest.* In other words, in order to phase reverse and subtract the competition, the circuit must have access to the waveform of only the competition before it is mixed with the speech signal of interest.

Figure 4 illustrates how noise-reduction headphones can meet the critical requirement of an independent version of the competition waveform. In this situation, the signal of interest travels via the electrical input to the diaphragm of the headphone. The noise arises from the environment around the listener. In the listener's ear canal is a mixture of the signal from the headphone (signal) and the competition that leaks in around the headphone cushion (noise). On the outside of the ear cup is a monitoring microphone that captures the noise signal directly before it is mixed with the signal in the listener's ear canal. After the spectrum of this sample of the noise is modified to account for factors such as the influence of the ear canal resonance, it is reversed and then added to the signal in the earphone. The end result is cancellation of a significant amount of the surrounding competition.

In audio sample 3, the effectiveness of phase cancellation is shown. In Figure 5, the upper panel presents the waveform of speech in quiet, while the middle panel shows the waveform of the speech mixed with complex, broadband noise (noise recorded in a café). Since the exact waveform of the noise was known (the sample was mixed offline), it was possible to reverse the noise sample in phase and then add it back into the speech-plus-noise signal. The result is shown in the lower panel of Figure 5. This addition of the phase-reversed noise essentially eliminated the background competition. However, it is rare to have an exact representation of the noise that is completely independent of the speech signal.

In headphone applications, such phase-cancellation approaches work best in the lower frequencies. In the higher frequencies, the shorter wavelengths make it difficult to precisely account for the high-frequency signal characteristics in the ear canal. However, for direct electronic applications, this system can be effective across the full bandwidth of speech as long as the competition is measured completely independent of the signal.

It should be clear from Figure 4 that the typical use of hearing aids does not meet the requirements for phase cancellation. The major problem is that the microphone of the hearing aid picks up a signal that is already a mixture of speech and noise. The circuitry does not have access to a "clean" noise signal, i.e., one independent of the speech signal.

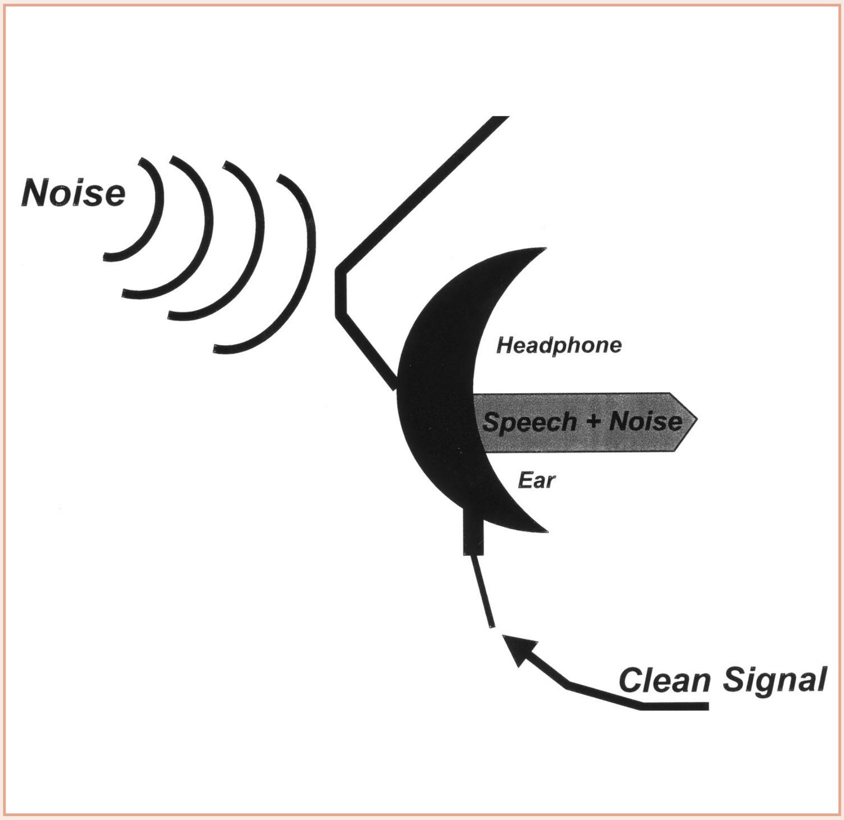

Figure 4. *A diagram of the conditions under which phase cancellation works in noise-reduction headphones. The mixed speech-plus-noise signal in the ear canal is based on speech being supplied to the headphone electronically and noise arriving acoustically from around the listener.*

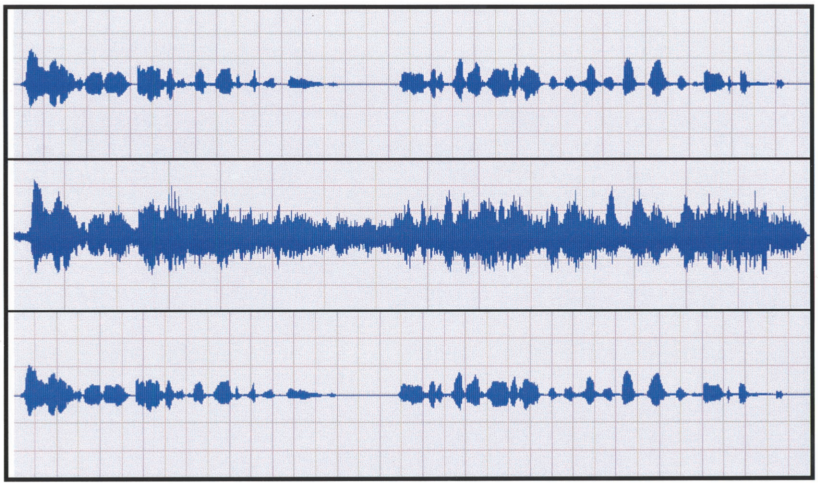

Figure 5. *Speech in quiet (upper panel), mixed with cafeteria noise (middle panel), and then after use of phase cancellation to remove the cafeteria noise (bottom panel).*

IMPROVING SPEECH UNDERSTANDING

There have also been DSP-based approaches that, rather than trying to reduce or eliminate noise, attempt to make speech more understandable against a given background of noise. In general, these approaches focus on identifying speech features or speech segments and then finding a way to improve their understandability.

Spectral enhancement

One such approach is spectral enhancement. Here is an example of how this works. Vowel formants tend to be robust against a background of noise and thus easily identifiable by a DSP algorithm. Once it has determined that a vowel is present, the algorithm will attempt to increase the contrast between the peaks and troughs in the spectrum.[1] The intention is to make the vowel more recognizable. Or, to give another example, the high-frequency noise of fricative consonants can be identified and then amplified above the rest of the spectrum.

Although this is an intriguing concept, its application in commercial products (including hearing aids) has been limited for a various reasons. First, nearly all reported studies of such strategies have shown small or non-existent gains in intelligibility.[2] At times, intelligibility actually decreases.

Further, the types of speech segments that such algorithms can identify most easily are the same segments that are already relatively well identified by the hearing-impaired listener (e.g., vowels). Patients generally have the most difficulty with the speech elements (e.g., unvoiced consonants) that are also the most difficult to identify via algorithm.

In addition, it is commonly observed that speech quality deteriorates when such artificial enhancements are applied. Apparently that's because isolating a single class of cues and manipulating them without making corresponding changes in natu-

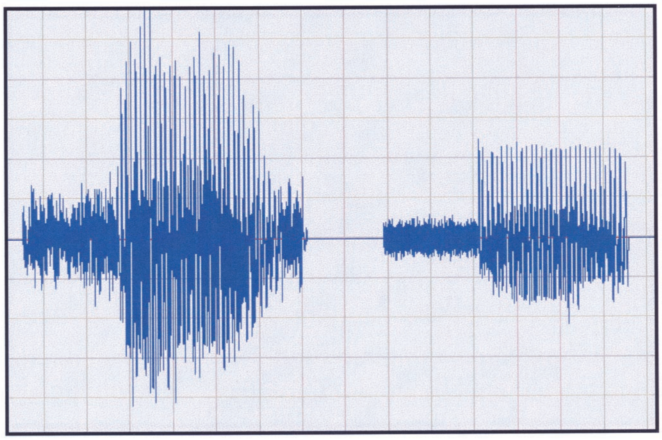

Figure 6. *The original, natural speech utterance "sale" presented in noise (left) and an artificially reconstructed version (right), as demonstrated in audio sample 4.*

rally associated cues throws the listener's perception off. Such speech does not sound natural.

Speech synthesis

A second general approach is speech synthesis or reconstruction. In this strategy, a DSP algorithm attempts to identify speech in a background of noise. Once the speech segments are identified, a new signal is synthesized carrying the same speech sounds. The advantage of this approach is that, as long as the original speech is accurately identified, the newly constructed signal will be free of background noise.

Audio sample 4 shows how speech can be reconstructed. In the left-hand section of Figure 6, the word "sale" is presented in a background of noise. In this hypothetical example, if an algorithm can determine that the first phoneme is /s/, then an artificial /s/ can be synthesized by taking white noise and filtering it to the spectral shape of a natural /s/. In addition, once the algorithm recognizes the presence of the diphone /al/, a stored version can be added in. The right-hand section of Figure 6 shows the reconstructed clean version. The goal is to create a noise-free version of speech by having the algorithm identify which speech sounds are present and then recreate them.

Speech synthesis must overcome serious challenges to become commercially viable. First, it is very difficult to devise an algorithm that can identify speech segments accurately against a background of noise or that can identify phonemes such as stops even in quiet.[3] Automatic speech-recognition systems (such as those used in speech-to-text or to "talk to your computer") need significant training for a particular talker and are still sensitive to background noise, especially competing speech.

Secondly, our ability to create natural-sounding synthesized speech is still very limited. Although words may sound natural in isolation, when they are strung together, the naturally produced suprasegmental features (pitch contours, stress patterns, pauses, etc.) are noticeably lacking. HJ

In the second and concluding segment of his article, Dr. Schum will discuss the use of noise-reduction schemes in hearing aids.

Donald J. Schum, PhD, CCC-A, is Vice-President, Audiology and Professional Relations, Oticon, Inc. Correspondence to Dr. Schum at DJS@Oticonus.com.

REFERENCES

1. Lyzenga J, Festen J, Houtgast T: A speech enhancement scheme incorporating spectral expansion evaluated with simulated hearing loss of frequency selectivity. *J Acoust Soc Am* 2002;112:1145-1157.
2. Stone M, Moore B: Spectral enhancements for people with sensorineural hearing impairments: Effects on speech intelligibility and quality. *J Rehab Res Dev* 1992;29:39-56.
3. Niyogi P, Sondhi M: Detecting stop consonants in continuous speech. *J Acoust Soc Am* 2002;111:1063-1076.