A common problem for people with a hearing impairment is the increased difficulty they encounter in communicating in noisy situations. Because the majority of people with a hearing impairment need to communicate in noisy situations at least some of the time, it is obvious that one design goal for hearing aids must be to improve their performance in noise.

Achieving this entails not only improving speech understanding in noise, but also increasing ease of listening and comfort and reducing listening strain and potential risk of additional hearing loss from excessive noise exposure. Although the Zeta Noise Blocker was the first commercial attempt to use digital signal processing (DSP) to improve speech perception in noise, the first commercially successful application of DSP in hearing aids for noise management was not available until 1996 when Widex introduced the Senso 100% digital ITE hearing aid.

The obvious way to overcome the negative effects of noise is to attenuate or eliminate the noise that reaches the hearing aid wearer’s ear. This can be accomplished easily if the environment is simply noise. The task becomes more difficult when speech is also present in the environment. In that case, the main challenge is to decrease the negative effects of noise without affecting, or minimally affecting, the intelligibility of speech and/or creating unpleasant acoustic artifacts. While it sounds simple and straightforward, there are many challenges that one must overcome before accomplishing this goal.

CHALLENGES TO IMPROVING PERFORMANCE IN NOISE

Challenge 1: Precise identification of wanted and unwanted sounds

Although we can agree that “noise” is something we do not want to hear, this definition is inadequate for the design of an artificial intelligence system that must respond appropriately to different signals. By that vague and subjective criterion, noise could be speech spoken by one person at a high volume, a continuous narrow band of noise at a high level, impulse signals, babble at a moderate level, low-level ambient noise and circuit noise, or someone playing a musical instrument poorly. Similarly, meaningful sounds such as speech, music, or warning signals that a listener does not want to hear could also be classified as noise.

Classifying all these sounds as noise would require an identification algorithm that is encompassing and yet specific enough to minimize the chance of an incorrect identification. One would need to make specific assumptions about “speech or wanted signal” and “noise or unwanted signal” so that their presence could be ascertained effectively and accurately. Thus the effectiveness of a noise-reduction algorithm is highly dependent on the applicability of the assumptions for the given situation. Only when the assumptions are valid will the identification of noise and speech be appropriate and the outcome of the noise-reduction algorithm desirable.

Earlier attempts at noise reduction in analog and programmable hearing aids assumed that noise was primarily low-frequency in nature. This assumption is not unreasonable because the long-term spectra of most noises are weighted heavily in the low frequency. Thus, simple fixed or adaptive high-pass filters were frequently employed as a noise-reduction approach. However, it is now recognized that such an assumption was too simplistic.

The use of digital signal processing techniques increases the precision with which “speech” and “noise” can be identified. One approach is to continuously analyze the short-term level distributions of the input signal (i.e., how frequently a brief segment of the signal occurs at a particular intensity interval). Figure 1A shows the 3-D level representation of recorded speech with children playing in the background. The X-axis represents the frequency regions corresponding to the 15 available channels in the Widex Senso Diva. The Y-axis represents the SPL at which the sound occurs, and the Z-axis (or vertical) represents the percentage of time that the SPL is within the indicated level bin.

One can see that in channel #1, the typical sound is recorded at about 40 dB SPL, while in channel #2 the most dominant level is about 75 dB SPL. As the channel number increases, the dominant sound pressure level decreases to about 35 dB SPL in channel #15. This is represented by the darker shaded area. On the other hand, in noise recorded from the workplace without any speech (Figure 1B), the dominant sound level in all 15 channels is about 50 dB SPL to 60 dB SPL. This indicates that the three-
dimensional representations of sounds (channel by intensity by percentage) for different sound environments are, like fingerprints, unique. Therefore, they could be useful in identification.

The level distribution of speech is significantly different from that of typical background noise. Typically, speech tends to spread across multiple intensity levels while noise tends to cluster around a single intensity region (Figure 4). This makes the use of level distribution a powerful tool for estimating the nature of the incoming signal. Other methods involve analyzing the modulation of the incoming signal and/or calculating the modulation spectrum, which can also be used to distinguish speech from noise.

**Challenge 2: Appropriate gain adjustment**

Once “noise” (or unwanted signals) and “speech” (or wanted signals) are identified, most noise-reduction systems attempt to adjust the “gain” in those channels and during those time intervals in which noise dominates over speech. The objectives are to minimize the direct masking caused by the noise in that channel, to minimize the potential spread of masking in adjacent channels, and to improve overall subjective comfort and sound quality.

The effectiveness of these noise-reduction systems on speech-recognition scores varies. In many cases, the magnitude of the changes has been small. One possible reason for such observation may be the “right” amount of gain adjustment that is applied.

Because the negative effects of noise, e.g., excessive noise exposure, upward spread of masking, and poorer speech recognition, occur primarily at high noise levels, one should design noise-reduction systems that exert the greatest effect at high input levels and less effect at lower input levels. This way, the negative effects of noise can be reduced with minimal effect on the audibility of the signal.

A concomitant consequence of gain reduction is that the overall output of the hearing aid decreases. When speech and noise occur in the same frequency channels and at the same time, the gain reduction for noise means gain reduction for speech as well. This reduces the perceived loudness of the listening situation and may also compromise intelligibility. Consequently, one important consideration in a noise-reduction algorithm is to compensate for such effects through the use of differential gain adjustment and loudness restoration. This is the rationale for the Speech Intensification System (SIS) used in the Senso+ hearing aids.

In the SIS algorithm, the estimated speech-to-noise ratios across channels are compared to further modify the amount of gain reduction in each channel. For the same speech-to-noise ratio (SNR), channels that convey more important speech information are attenuated less than those that convey less speech information. The goal is to increase listening comfort while reducing the masking effect of noise with minimal negative effect on overall loudness and intelligibility. This approach has resulted in significant improvement in speech intelligibility over the original Senso noise-reduction algorithm where the SIS was not available.

**Challenge 3: Determining the number of channels**

The efficacy of today’s noise-reduction algorithms may be improved with a finer identification of speech and noise signals. We can use the spectrogram of the sentence “Will you please confirm government policy?” for illustration (Figure 2a).

The spectrograph displays the frequency (Y-axis) and time (X-axis) information of the acoustic signal with the intensity shown by the darkness of the display. For example, Figure 2a clearly shows the fundamental frequency (dark, shaded area on the bottom of the figure) and the distribution of energy across different frequency regions, as well as the changes in formant frequency with time.

This analysis was done with a filter bandwidth of 50 Hz. If the spectrogram had...
been performed with many fewer analysis channels or with a broader bandwidth, the finer details of speech, e.g., the transition of formants, would not be easily visible. This is illustrated in Figure 2b, which shows the spectrogram performed with a bandwidth of 1500 Hz on the left and one with a bandwidth of 50 Hz on the right. Information on the formant transition is almost unrecognizable from the display on the left (with a bandwidth of 1500 Hz). This suggests that the analysis performed with few channels may not reflect all the information available in the acoustic input signal.

The resolution of the speech and noise analyses can be improved by narrowing the channel bandwidths and providing more independent filter channels. By doing this, one can also focus the effect of the noise reduction on just the “noise” frequency without affecting other frequency regions. Intuitively, it may seem desirable to have as many filter channels as possible for accurate signal identification and noise reduction. However, two potentially negative consequences result from an increase in the number of channels.

First, filters with a narrower bandwidth impose a longer group delay than filters with a broader bandwidth. This means that the processed signal reaching the eardrum will be delayed relative to the acoustic signal that reaches the eardrum directly (e.g., via the vent). Although it is generally assumed that small delays are harmless, recent research has suggested that delays as short as 5 to 10 ms may have a negative impact on sound quality (own voice sounds hollow or echoic) for listeners with near-normal hearing at low frequencies.16

Second, a large number of filter channels imposes a higher risk of spectral smearing. Smearing occurs when the filter channels are allowed to adjust their gain independently. The consequence is that the intensity contrasts across filter channels could be reduced, leading to poorer sound quality and speech intelligibility.17

Challenge 4: Optimizing the speed of noise reduction

The speed at which a filter changes its characteristics can have a significant impact on its effectiveness in noise management. Conventional analog hearing aids are typically designed with filters having fixed characteristics. That is, the same filter characteristics are used for all signals at all times and all input levels.

One of the many advantages of using digital signal processing in hearing aids is that the filters can be given characteristics that vary depending upon the incoming signals. For example, in cases where the noise and speech spectra do not overlap, where noise is confined to a narrow frequency region, or where the environment is primarily noise, the use of filters with a slow regulation time can result in significant reduction of the noise spectrum and improved communication. This is because noise typically varies more slowly than speech in its short-term spectra.

However, when speech and noise are both present in the same frequency channel, the use of fixed or slow response characteristics may not improve the effective SNR. The slow-varying regulation may allow the filter to follow the slow short-term variations of noise but will be unable to follow the fast short-term variations of the speech spectrum. As a result, both speech and noise within a filter channel would be affected to the same extent. This may explain why studies failed to demonstrate the SNR advantage of some noise-reduction algorithms.18,19

One may be tempted to think that the solution to improving the efficacy of DSP noise-reduction algorithms is simply to increase the regulation speed of filters so that they can follow the more rapid variations of the speech spectrum (i.e., fast-acting noise reduction). Unfortunately, increasing the speed of regulation tends to create a number of side effects such as musical noise, frequency smearing, temporal smearing, and a noisier sound image.

An examination of the detailed spectrogram shown in Figure 2a reveals that within any frequency region there are times when there is no or minimal activity, i.e., pauses. If one assumes that any activity during pauses is noise, one can apply gain reduction during these pauses to significantly improve speech intelligibility and listening comfort even when the noise has a speech-shaped spectrum. Thus, any noise-reduction filter must be able to vary its characteristics over time to achieve optimal performance.
On the other hand, one must be careful not to create any spectral or temporal smearing because hearing-impaired people, especially those with a severe-to-profound loss, rely on the waveform envelope for speech identification.20

**STRATEGIES USED IN THE DIVA NOISE-REDUCTION SYSTEM**

The important variables discussed here have been considered in the design of the noise-reduction system in the new Senso Diva hearing aid. Figure 3 shows a functional block diagram of its noise-reduction algorithm.

**Preserving audibility while minimizing negative noise effects**

Figure 3 shows that each processing channel includes an independent noise-reduction (NR) module coupled to a compression module (AGC). The AGC module adjusts the output of the channel to a level that is dependent on the residual dynamic range of the wearer. This module is placed prior to the noise-reduction module so that only a high-level output (resulting from a high-level input) is “treated” by the noise-reduction algorithm.

This is appropriate because the negative effects of noise are most harmful at high levels. Reducing the high output lessens the risk of over-amplification and increases listening comfort while preserving audibility of the signal. On the other hand, if the noise-reduction algorithm also reduces the output of a low input signal, audibility for soft sounds, which may include distant warning signals, would be affected without improving comfort.

**Method of noise estimation**

One statistical algorithm used by the Diva to estimate the nature of the incoming signal is the level-distribution function (or histogram), similar to those displayed in Figure 1. Simply, this function displays the instances when a particular intensity level occurs within a measurement time window (interval).

For example, Figure 4a shows the level distribution of noise recorded inside an automobile within a 125-ms time window. The X-axis shows the SPL at which the signal segment occurs, and the Y-axis shows the percentage of times that a particular SPL is recorded during a measurement period of 120 seconds. The unimodal appearance of the distribution suggests that the typical SPL occurs about 70 dB SPL.

Figure 4b shows the level distribution of female speech. In this case, one notes that the output at 25 dB SPL occurs about 2% of the time while the output at 65 dB SPL occurs about 10% of the time. The distribution is bi-modal. The “peak” at the low levels indicates that speech contains a large portion of weak segments, including pauses between sentences and words, and closure of stop consonants. The “peak” at the high levels represents intense speech sounds such as vowels and intense consonants. By continuously monitoring the distribution of short-term intensity levels over a period of 10-15 seconds and by extracting the frequency and level information of noises that occur during pauses, one can estimate the speech and the noise SPL in each channel.

Each noise-reduction unit provides a running statistical analysis of the SNR of the input signal in the frequency range covered by the specific channel. No gain adjustment is made if the SNR in a channel is favorable (e.g., >+20 dB). Otherwise, noise reduction occurs in a graded manner. In general, the amount of gain reduction increases as the SNR worsens and as the overall level increases. That is, maximum noise reduction occurs when the SNR is poor and when the input level is high. As much as 14-dB gain reduction will occur in the worst listening condition. An expanded view of the noise-reduction (NR) module is included in Figure 5. The two analysis modules that provide the running estimates of the speech and noise levels are designated “speech analysis” and “noise analysis,” respectively. The outputs from both analysis modules are used to estimate the SNR in each channel.

**Increasing precision while minimizing artifacts**

To increase the analyzing power of filters, their bandwidths are designed to be only one-third of an octave wide throughout the main speech frequency range (400 Hz to 4000 Hz). This is close to the critical bandwidth of the auditory system. This approach results in a total of 15 independent processing channels in the Diva.

To further increase the precision of gain adjustment, each filter has a slope that can be as steep as 50 dB/octave. To minimize the group delay inherent in filters with such a narrow bandwidth, infinite impulse response (IIR) minimum delay filters are used. In addition, a high sampling frequency (32 kHz) is used to further minimize group delay.21

In the Diva, the group delay is typically 2 ms. An anti-smearing unit is included prior to the AGC module to maintain the temporal and spectral contrasts of the input signal while increasing the resolution of the NR algorithm.

**Preserving loudness and enhancing intelligibility**

Unlike noise-reduction systems that apply graded gain reduction based solely on the results of the SNR analysis, the Diva’s system has an additional module to enhance
Improving hearing aid performance in noise

The intelligibility of speech in noise. This is achieved through the Speech Intensification System, in which information on the speech importance weighting and the loudness contribution of each frequency channel is stored. Once the estimated SNR information from each channel is routed to this block, a cross-channel analysis is performed to further validate the estimated SNR in each channel.

Afterwards, the speech importance weighting and the loudness weighting of each channel are applied to each channel so that channels with the higher speech information experience less gain reduction (or noise reduction) than channels with lower speech information. Information on the loudness contribution is useful in ensuring that the loudness of the overall output is significantly reduced while the loudness of speech is maintained.

These considerations result in gain reassignment across channels from the original settings. The reassignment is based on the degree of hearing loss and input level, as well as the estimated SNR in that channel alone. The goal of the SIS module is to ensure that the processed signal containing both speech and noise (or speech alone) has the most comfort, most natural loudness, and maximum intelligibility. In the case of noise alone, the noise-reduction algorithm is allowed to operate with maximum attenuation. Finally, the output from all the channels is summed to yield the ultimate output.

Optimal speed of time-varying filter

The objective of the SIS is to define the characteristics of the noise-reduction filters (e.g., amount of attenuation, slope). Because of the varying nature of the input signals, noise estimates and speech estimates are constantly updated at 32 kHz in order to define the optimal characteristics of the time-varying filter at any time. To avoid potential artifacts such as musical noise, the rate at which the filter varies its characteristics is further empirically optimized. In general, a fast rate is used where speech in noise is detected and a slower rate is used where primarily noise is detected.

Figure 6 shows how the frequency gain characteristic changes over a 25-second period in (a) party noise and (b) car noise. Note the changing characteristics over time in both noise environments.

Realistic expectation for the algorithm

Figure 7 shows the input and output waveforms of several signals processed by Diva’s noise-reduction algorithm. Figure 7a shows those of a signal with speech alone. Note the similarity between the output and the input. Figure 7b shows a speech signal embedded in a noise background. The processed output shows much less noise, and clearer demarcation of the speech signal. This may improve sound quality, increase listening comfort, and even enhance speech intelligibility. In the last case (Figure 7c), which contains only noise input, one sees a substantial reduction in the magnitude of the output waveform.

IMPACT ON WEARERS

The noise-reduction algorithm used in the Senso Diva can be expected to improve effective signal-to-noise ratio in many situations. However, the magnitude may be small if the wearer is not negatively affected by the upward spread of masking or has poor speech-recognition potential (i.e., poor speech understanding in quiet).

Furthermore, despite its sophisticated noise-reduction algorithm, the hearing aid is not “fool-proof.” For example, consider the following situation where “noise” is actually “speech”: Two people are talking behind a listener who is attempting to listen to another speaker in front of him. The speakers behind could be considered a noise source. Yet, the noise-reduction algorithm may conclude that the signal from behind is speech and therefore not initiate any noise reduction.
Listening to music is another situation in which the algorithm may not be helpful. The hearing aid might identify the music as noise and reduce its gain.

Examples like these illustrate why wearers of the hearing aid need to be counseled properly and given realistic expectations. Having the option to deactivate the noise-reduction algorithm in certain situations may be desirable.

Despite such caveats, we believe this algorithm offers the possibility of improved speech intelligibility, increased listening comfort, and reduced listening strain. As a result, the total benefit for the hearing aid wearer can be substantial.

REFERENCES


