Challenges and Some Solutions for Understanding Speech in Noise

There are two difficult hearing instrument situations created by competing noise. One involves internal hearing instrument noise, the other involves external noise. Internal noise is electronic noise generated inside the hearing instrument. This noise is present whether or not there is any sound arriving at the microphone. If this noise occurs at a high enough level, it may be distracting for the listener, or it may interfere with speech comprehension due to masking of low level sounds or the creation of distortion.

Typically, however, when hearing instrument users complain about noise, they are referring to noise external to the hearing instrument and the related difficulties in understanding in noisy environments. The problem of separating speech from external competing ambient noise with current hearing instrument technology is a difficult one. The term noise is generic and encompasses a variety of difficult listening situations. For example, a listener may be trying to understand one particular speaker against a background of many talkers, such as at a social gathering. Alternately, the listener may be trying to identify one or several talkers amid a background of surrounding noise, such as the clatter of dishes, doors slamming, other talkers and background music in a busy restaurant. A third situation may be that of trying to understand speech amid a background of non-voice competing low frequency sound, such as following a conversation on a busy street, in a car or in an airplane. Each of these scenarios presents a different speech extraction problem because the competing background noise is different in each case. In order to produce a perfect solution for extracting intelligible speech from surrounding noise, the hearing instrument must be able to identify the desired speech while suppressing the background noise, whether the undesired noise is competing multiple talkers or interfering environmental artifacts.

The most challenging task is for the hearing instrument to identify and extract one desired talker in a situation with multiple talkers. This solution is also the one most often desired. Because the voices of most talkers are similar in spectrum and amplitude, the ideal hearing instrument would be able to identify the individual speech characteristics of a particular desired talker and distinguish it from others, so that the words and syllables of the desired talker are not interspersed and confused with those of other talkers. This separation must be achieved while maintaining a high quality of processed sound, so that the desired speech is audible and readily recognizable. Obviously, this is a very difficult challenge for any hearing instrument or processing algorithm, even though this task is easily accomplished by a normal functioning human auditory system.

Part of the problem from the hearing instrument design viewpoint is that the hearing aid can only differentiate sounds based on three fundamental physical characteristics.

1. The amplitude of the sound. This is the loudness of the sound at the hearing instrument microphone. The amplitude of a single speaker in a noisy party situation is often the same as (or may be lower than) the surrounding noise. This reduced difference in levels between the speaker and the noise is what causes a speaker to instinctively raise his/her voice above normal levels in order to be clearly understood.

Jeremy Agnew, PhD, is director of product development at Starkey Laboratories, Eden Prairie, MN.
is an effect called the Lombard voice reflex. If the amplitude of the desired speech is higher than that of the undesired noise, it may be possible to extract intelligible speech by electronically expanding this difference in amplitude between the two. If the desired speech is lower in amplitude than the surrounding noise and has the same spectrum, then it is difficult to separate the two.

2. The frequency of the sound. For speech, this usually includes multiple frequencies at any instant of time (recalling from Fourier analysis that every sound is composed of the sum of a series of sine waves). In most listening situations with competing babble, the long-term spectrum of the desired voice is the same as that of undesired competing voices, thus providing no basis for extracting speech via long-term spectral differences. The instantaneous spectra of the two may differ; however, as the number of competing voices increases, the problem of extraction becomes more difficult. Early attempts to utilize spectral differences were developed as the so-called automatic signal processing (ASP) hearing instruments, such as those described by Ono', Kates', and Staab and Nunley. The intended concept was that an automatic reduction in the amplification of low frequency energy would reduce any upward spread of masking and thus improve intelligibility. Though initial hopes were high for improvements in speech intelligibility using this processing methodology, later studies showed that intelligibility was not improved. It is now accepted that the attenuation of low frequency amplification in background noise will improve listening comfort and lessen fatigue by reducing the level of annoying low frequencies, but that this type of fitting strategy will not improve intelligibility. However, while not the initially-desired result, a lowered level of listening fatigue can still be important for user satisfaction.

3. The duration of the sound. This relates to when and how long the sound exists in time. Obviously, if the desired speech and undesired competing sound do not occur at the same time, then there will be no problem in understanding the speech. However, a more likely scenario is that the speech and the undesired background noise are received at the microphone of a hearing instrument with identical long-term frequency responses and amplitude at the same time. If this occurs, as may happen when competing talkers are present, there are no fundamental long-term signal differences that can be used to easily separate the speech from the noise. There may, however, be ways to use short-term differences in the spectra, as will be touched on below.

### Signal-to-Noise Ratio (SNR)

A typical frustrating scenario for a hearing instrument wearer is that of trying to understand speech in a noisy restaurant. The listener’s immediate reaction is to turn up the gain control in order to improve the audibility of the speech. Unfortunately, what also happens is that amplification of the undesired noise occurs at the same time. While the listener perceives the speech as being louder, the noise is also proportionately louder. Thus, the relative difference between the speech and the noise has not increased and intelligibility has not been improved.

One key to improving speech intelligibility in noise is to improve the amount by which the speech (i.e., the desired signal) exceeds the background noise—in other words, improving the signal-to-noise ratio (SNR). The SNR is expressed in decibels, with a positive SNR if the signal is higher than the noise. A negative SNR denotes that the noise is higher than the desired signal. A favorable SNR situation would be speech occurring at 70 dB SPL in a quiet room with an ambient noise of 40 dB SPL, thus giving an SNR of +30 dB. A much more difficult listening situation occurs with loud speech at 85 dB SPL with a background babble of 83 dB SPL, or an SNR of only +2 dB. (See Fig. 1 for a more in-depth discussion.)

Individuals with normal hearing can usually still understand speech with an SNR of -6 to -10 dB. By contrast, individuals with a hearing impairment may require 6 to 10 dB of additional SNR above that required by an individual with normal hearing. Further discussion on the amount of SNR required for different degrees of hearing loss may be found in Kil-lion.

### Practical Hearing Instrument Solutions

While the perfect hearing instrument solution to extract transparently one voice cleanly and clearly from an undesired background of multiple talkers by a hearing instrument is not readily apparent at this time, there are useful fitting techniques for improving the listening situation for a hearing aid user. The following offers a summary and commentary on the technology of today and of the imminent future. (See Fig. 2 on page 6 for a more in-depth discussion.)

#### Wider bandwidth

In general, the wider the audible bandwidth of the hearing instrument, the higher the intelligibility scores. This finding relates to the audibility of speech cues and the increase in available bandwidth in modern hearing instruments compared to those of 30 years ago. As more high frequency speech cues are presented to the listener, the more speech will be understood. While this may seem like an obvious solution, it is not always used.
Skinner and Miller' tested listeners with a hearing impairment in quiet and in noise using a master hearing aid that allowed variation of the bandwidth of the speech presented to the subjects. For all subjects, the highest scores were obtained with the hearing instrument set to deliver the widest bandwidth. These authors further suggested that a reduction of low frequency amplification may be detrimental to speech understanding, because this results in some low frequency cues becoming inadequate or inaudible. Vershure and van Benthen' also noted that appropriate high frequency amplification improves performance for listening in noise.

**Binaural hearing instrument fittings.** Decoding of speech cues is performed by the auditory cortex. To help perform this function, the brain depends on an input from each ear. The ability to correctly perceive the location of sound in space depends in part on the difference in the amplitude and arrival time of sound at the two ears and on the alteration of the spectrum of the incoming sound by filtering effects of the pinna. The perception of the location of sounds outside the head (externalization) is also related to the correct presentation of binaural auditory cues. See Batteau', Rodgers' and Blauert' for further details. Binaural processing of these cues has been shown to be a significant factor in increasing intelligibility.'

One advantage enjoyed by listeners with normal hearing in noisy situations is the ability to take advantage of the difference in spatial location that often occurs between a desired speaker and undesired noise. If spatial separation is present, the normal functioning of the central auditory system allows suppression of undesired background noise and the ability to concentrate on desired speech. Dirks and Wilson' and Ter-Horst et al.' showed that intelligibility scores were greater when sources of speech and noise were spatially separated. Chappell et al.' found that 15-of-18 normal-hearing subjects perceived a greater distinction among voices when using binaural listening than with monaural listening, due to the separation of the sounds in space. These listeners showed an average improvement of 20% in intelligibility under binaural conditions. The effectiveness of binaural listening over monaural listening can be easily demonstrated by plugging one ear with a finger while trying to understand speech from a nearby talker in a noisy restaurant.

An appropriate fitting of binaural hearing instruments will improve the ability to understand speech in noise by preserving and presenting binaural cues to the central auditory system. Typically, binaural fitting in noise produces intelligibility scores which are higher compared to monaural listening at threshold levels. At suprathreshold levels, the improvement can be 6-8 dB at moderate sensation levels.' Thus, one simple way of improving speech intelligibility is to use a binaural fitting in appropriate situations (see Byrne' for further details).

**Incorporation of individual head-related transfer functions (HRTF).** Another fitting strategy under investigation is the use of the head related transfer function (HRTF) in signal processing. The HRTF produces a modification of the spectrum of sound perceived by a listener and is the result of spectral filtering due to the interactions of sound with the pinna, combined to a lesser extent with reflections and diffraction from the head and upper torso. The convolutions of the pinna have been shown to be responsible for the externalized quality of sound images.

HRTFs differ between individuals and provide each listener with a unique way of perceiving the locations of sounds in space. The difference between the peaks and valleys in the HRTF typically extends over a range of 50 dB. The location of the peaks and valleys may vary over a span of 1000 Hz between individuals, with wide variations of the levels between. Alteration of the perception of sound modified by the individual's HRTF that has been developed over a lifetime has been theorized to disrupt localization cues.

**Completely-in-the-Canal (CIC) hearing instrument fittings.** CIC hearing instruments offer many benefits, including high cosmetic acceptability and a clear sound. One of the important benefits of CIC hearing instruments may be that the microphone placement of a CIC hearing aid within the entrance to the ear canal leaves the convolutions of the pinna exposed, and may allow spectral shaping of the sound in a manner consistent with the condition without a hearing instrument before it enters the microphone.' This type of fitting provides high frequency audibility and maintains the individual's HRTF. Some binaural CIC hearing instrument wearers have described the availability of high frequency cues provided by CIC hearing instruments over ITC fittings. Some CIC hearing aid users also state that the sounds are in the right place,' indicating that they perceive sounds to be external to the head.' This possible explanation may be that a CIC fitting maintains concha-related resonances found between 4000 Hz and 5000 Hz.' Fortune' has also described the availability of high frequency cues provided by CIC hearing instruments, particularly between 5000-6000 Hz.

**Low distortion.** Though distorted speech remains highly intelligible for tests of speech reception in quiet with normal-hearing listeners, intelligibility is reduced with an ambient noise background. By definition, distortion in a hearing instrument creates undesired elements in the sound delivered to the hearing aid wearer that are not present in the input sound. Since these additional products are frequencies that are not present in the input signal, distortion creates additional confusing cues that effectively generate masking noise and disrupt speech comprehension. In addition, electronic noise generated inside the hearing instrument may either directly mask

---

**Fig. 2: Some Solutions for Noise**

<table>
<thead>
<tr>
<th>SOLUTION</th>
<th>BASIS FOR SOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wide Bandwidth</td>
<td>Adds speech cues in high frequencies.</td>
</tr>
<tr>
<td>Binaural Fitting</td>
<td>Makes maximum use of residual binaural effects.</td>
</tr>
<tr>
<td>HRTF</td>
<td>Maintains binaural cues to auditory system.</td>
</tr>
<tr>
<td>CIC Hearing Instruments</td>
<td>Maintain high frequency audibility and openness of pinna.</td>
</tr>
<tr>
<td>Maintain Low Distortion</td>
<td>Stops generation of unwanted frequencies that mask speech cues.</td>
</tr>
<tr>
<td>Variable-Release Compression</td>
<td>Maintains syllabic separation.</td>
</tr>
<tr>
<td>Direction instruments</td>
<td>Improves SNR in desired direction.</td>
</tr>
<tr>
<td>Beamforming Instruments</td>
<td>Produces narrow beam to allow focus on desired sound source.</td>
</tr>
<tr>
<td>Speech Processing Algorithms</td>
<td>Uses various DSP solutions to extract speech from background.</td>
</tr>
</tbody>
</table>

---

6 HIGH PERFORMANCE HEARING SOLUTIONS VOL 3 • JANUARY 1999
out quiet speech sounds or may react with low level sounds to create intermodulation distortion.

Both Hawkins and Naidoe and Tede? have hypothesized that some of the difficulties mentioned by hearing instrument users in noisy situations may be due to saturation-induced distortion. The effectiveness of two-microphone arrays has been described by Valente et al., who tested a commercially-available hearing instrument that showed an average improvement in SNR of 7.4-8.5 dB in the directional condition in comparison to the omnidirectional mode, when tested with speech from the front and competing noise from the rear. Agnew and Block, using a different design of a two-microphone array, showed an average improvement of 7.5 dB. These results were obtained under ideal test conditions, thus the increases in SNR...

Fig 3: Example of front and rear frequency responses that may be observed with modern directional hearing instruments.

Hawkins and Vacullo showed an improvement of 3-4 dB in SNR when using a directional hearing instrument in difficult listening situations. Dillon and Macrae found a similar advantage for directional hearing instruments over non-directional hearing instruments. Depending on the performance-intensity (PI) function for the particular test material used, an increase of 2-3 dB in SIVR may result in an improvement of 20% to 30% in intelligibility score.

Traditional directional hearing instruments use one directional microphone with two sound inlet ports. Extensive information describing two-port directional microphones utilized in hearing instruments may be found elsewhere. Newer directional hearing instruments use two omnidirectional microphones with electronic signal processing to produce a variety of directional patterns. The ability to switch from directional to omnidirectional and to vary the front-to-rear attenuation may be included.

The effectiveness of two-microphone arrays has been described by Valente et al., who tested a commercially-available hearing instrument that showed an average improvement in SNR of 7.4-8.5 dB in the directional condition in comparison to the omnidirectional mode, when tested with speech from the front and competing noise from the rear. Agnew and Block, using a different design of a two-microphone array, showed an average improvement of 7.5 dB. These results were obtained under ideal test conditions, thus the increases in SNR...
reported in these studies will be tempered in situations where the competing noise may arrive from other directions. Fig. 3 shows typical front and rear frequency responses that may be obtained from modern directional hearing instruments under ideal conditions. The difference between the two curves is the attenuation for sounds from the rear direction.

Positive benefits from the use of hearing instruments using a two-microphone directional array have been reported by Kuk. Newer hearing aid microphones with directional patterns narrower than the microphones that have been available for traditional ear-level directional hearing instruments, such as the D-MIC*, have been introduced.

10. Beamforming hearing instruments. By incorporating two or more microphones into the hearing instrument, increased directionality may be achieved. Microphone arrays developed with different concepts have been described by Peterson et al. and Soede et al.*

Extremely narrow reception patterns, or “beams,” may be achieved that are used by the listener to “focus in” on a particular point in space. This allows high rejection of competing noise from the side and rear directions, resulting in an improved SNR from the front direction. This type of directional hearing instrument system currently does not have universal appeal because of a larger configuration required by the wide microphone spacing in order to achieve effective beam-forming.

One practical example of a broadside array using multiple microphones that is currently undergoing clinical testing has been described by Widrow. In this configuration, an array of six microphones is worn on the chest in a device shaped like a necklace (for an illustration of the device, see pg. 00). Though this device is larger than traditional hearing instruments, the basic concept of the strategy is effective. Future development of sub-miniature microphones will reduce the size and configuration of these types of devices into forms more acceptable for mainstream hearing instrument wearers.

Sophisticated DSP algorithms are also being used to achieve sharply-focused patterns of sound reception with only two microphones. This technique uses a single microphone in each ear and has been adapted for acoustic use in hearing instruments from technology used in high-directional sonar and radar systems. Such techniques can provide improvements of 20 dB in front-to-back SNR in rooms with low reverberation. Experimental binaural devices using these techniques have been tested with hearing-impaired users.*

11. Speech processing algorithms. Attempts to separate the components of speech through signal processing have had mixed results. One challenge for the hearing instrument designer is to incorporate a sufficient amount of processing power within the limitations of a cosmetically-appealing hearing instrument and the power capabilities of currently-available hearing aid batteries. On-going developments in algorithms and integrated circuit engineering continue to yield promising results.

One promising approach for DSP algorithms is to develop processing of the short-term components of speech. For example, one processing scheme splits the incoming sound into several bands and analyzes the modulation frequency in each band. If this frequency is within the modulation frequency range of speech, the incoming sound is processed without alteration. If the modulation frequency is not within this range, the algorithm assumes that the incoming sound is noise and produces attenuation of the band containing that frequency. This scheme, therefore, tends to attenuate the noise component of incoming sound while maintaining the speech. Other processing schemes look for statistical differences in the properties of speech and noise, and perform different processing depending on which is estimated to be present.

If the spectrum of desired speech is different than that of undesired noise, it may be possible to enhance the SNR by other filtering techniques. If the noise occurs in a narrow band, or bands, it may often be successfully removed by filtering out that band*, while still retaining enough speech cues in the remaining spectrum to improve intelligibility.

Another signal processing scheme uses close and far microphones in conjunction with adaptive DSP noise-cancelation algorithms.* The far microphone is located away from the talker and picks up only noise, while the close microphone picks up both speech and noise. DSP processing is then used to subtract the two signals and extract the speech. This method is particularly useful in environments with stationary noise, such as in tanks and helicopters, but has been difficult to adapt for practical ear-level hearing aids because the two microphones must be spaced far enough apart for the algorithm to be effective.

Another practical aspect of the problem is that effective signal processing must be performed very rapidly. In order to provide meaningful two-way conversation, the processing of the desired speech must be performed in real-time. While it is possible to post-process recorded speech and noise then extract intelligible speech, many algorithms that successfully perform this are so complex that they do not operate in real-time. For practical use, a delay of more than about 20 milliseconds between seeing words spoken and hearing the speech may cause confusion for the listener during a conversation.

Further reviews of different speech extraction and enhancement algorithms may be found in Williamson & Punch* and Murray & Hansen.*

Conclusions
While there is continued research to solve the problem faced by hearing instrument users in noisy situations, the “perfect” transparent solution has not yet been found. However, as this paper describes, there are practical solutions that may be implemented with current technology. Extended bandwidth, maintenance of binaural hearing cues, fitting instruments with low distortion, and directional microphones all extend the availability of hearing cues for the listener and can result in improvements for hearing in noise.

References

Correspondence can be addressed to HR or Jeremy Agnew, PhD, Starkey Laboratories, Inc., 11 110 Ektlon Dr., Unit E, Colorado Springs, CO 80907; e-mailjerry_agnew@starkey.com.

VOL 3 • JANUARY 1999
HIGH PERFORMANCE HEARING SOLUTIONS