Introduction

In audiology, there is much confusion about whether speech perception declines with age, and if so, what is the nature of that decline. This article will review the age-related declines in speech perception, review the evidence that separates decline due to hearing loss from decline due to another age-related process, discuss older listeners’ need for acoustic redundancy in the speech signal, and describe the impact of these findings for hearing instrument amplification.

Challenges to Speech Perception

Older listeners often have a dual challenge when it comes to listening to speech: both hearing loss and higher-level age-related degradations can interfere with their clear reception of the speech signal. Clinically, we deal with both of these factors in the same person at the same time, so it might seem like it’s just an academic question to know how much speech recognition deficit is due to hearing loss and how much is due to higher-level processes.

However, in a clinical situation, it is helpful to be able to separate out factors of aging and hearing loss. If a 35 year old and an 85 year old present with the same audiogram, we need to know what we should be able to expect for differences in speech recognition, whether this has any implications for the type of hearing instrument processing we choose, and whether our expectations of outcomes and benefit should be different, at least at the level of speech recognition. Thus, it is important to be able to separate out declines in speech recognition due to hearing loss (i.e., reduced audibility) from declines in speech recognition due to higher-level age-related processes, such as declines in the central auditory system or cognitive changes.

Recognition decline associated with peripheral hearing loss can best be treated with amplification that provides audibility of signals. Declines associated with either central auditory or cognitive mechanisms will not necessarily be helped by amplification processing (although inappropriate processing could make the listening situation more difficult by introducing additional signal distortion). At the cognitive level, declines are best remedied with compensation strategies, such as use of context, asking the talker to speak more slowly and with pauses, and relying on visual cues provided by the talker. At the central auditory level, remediation may involve listening training (Tremblay, Kraus, McGee, Ponton and Otis 2001) or amplification systems that improve the signal to noise ratio, such as personal FM systems.

Simple Speech

It has been shown several times in the research literature that once hearing loss is equated for young and older listeners, speech recognition on a simple speech task remains constant with age (Cheesman, Hepburn, Armitage and Marshall 1995; Divenyi and Haupt 1997a, 1997b, 1997c; Gelfand, Piper and Silman 1986; Souza 2000; Souza and Kitch 2001). Now, what do we mean by simple speech? A good definition is speech in quiet that has not been distorted or degraded. The definition can be expanded to include speech in a background of unmodulated, meaningless noise, such as white or broadband noise.

Contact Information: Lorienne Jenstad, Ph.D, Assistant Professor, School of Audiology and Speech Sciences, University of British Columbia, 5804 Fairview Avenue, Vancouver, BC V6T 1Z3, Canada. Email: ljenstad@audiospeech.ubc.ca. Phone: 604-822-4716.
For simple speech, reductions in recognition for many older listeners can be explained by worse hearing thresholds in older listeners. This finding has been demonstrated for vowel perception in quiet (Nabelek 1988), nonsense syllable perception (Helfer and Huntley 1991), and word recognition in quiet (Dubno, Lee, Matthews and Mills 1997). Importantly, this relationship holds true for most speech that has not been temporally distorted (Humes 1996).

The belief that recognition of simple speech is unaffected by age-related processes beyond hearing loss has not always been widely accepted, because there is often a confound between age and hearing loss; that is, it is difficult to find younger and older listeners with the same degree of hearing loss. Usually the degree of hearing loss increases with increasing age, even for listeners with so-called normal hearing. Thus, when results differ between the younger and older groups, it’s difficult to know how much of the effect is due to hearing thresholds and how much is due to other aging-related processes.

Individual Differences

Not all speech understanding in older listeners is explainable by the simple audibility model (i.e., that their degree of hearing loss reduces audibility of speech). Some older individuals have difficulties in quiet speech understanding to an extent not explained by their hearing loss. Defining older listeners has been the subject of some debate. The exact age group used to test older listeners is widely variable across studies, sometimes including all listeners over a specific age as a single older group, or limiting subjects to an upper age limit (of 75 years and below, for example). Greater reductions in speech recognition than would be predicted by threshold change are more likely seen in old-old listeners (over 75 years) than in young-old listeners (age 65 to 75 years) (Magnusson 1996; Sherbecoe and Studebaker 2003). A mean age difference of 10 years can have a large effect on the results of different studies (as noted in Souza 2000). This difference is independent of the amount of hearing loss, shown by Humes and Christopherson (1991) where their old-old subjects performed worse for filtered and reverberant speech than the young-old, despite equivalent hearing losses.

Complex Speech

Age, independent of peripheral hearing loss, becomes a more important contributor to speech recognition for complex listening tasks (Schum, Matthews and Lee 1991). Older adults, particularly old-old listeners, tend to have more difficulty than younger listeners in noise, and even more difficulty when the background noise is temporally modulated (Dubno, Horwitz and Ahlstrom 2002). Other complex listening tasks that reduce speech recognition performance are speech that has been distorted by reverberation or other temporal manipulation (Helfer and Huntley 1991; Humes 1996; Nabelek 1988). While it is difficult to separate out the effects of aging from the effects of greater hearing loss found in older adults, studies that have employed careful controls for hearing loss have shown that increasing age is a predictor of reduced speech recognition under these distorted or degraded listening conditions (e.g., Studebaker, Sherbecoe, McDaniel and Gray 1997). Figure 1 shows hypothetical data for a common pattern found for speech recognition performance: an interaction between age of listener and complexity of listening task. For simple listening tasks, there is usually no difference in speech recognition performance between younger and older listeners when the two groups are matched for hearing. As the complexity of the task increases, performance is reduced for younger listeners and even more for older listeners. Note that although this figure is not plotting actual data, this pattern of performance is quite commonly found.

Figure 1. Hypothetical data illustrating a common finding of an interaction between listening condition (either simple or complex) and age of listener on performance on a speech recognition task. When younger and older listeners are matched for degree of hearing loss, there is usually little to no effect of age on speech recognition for simple listening tasks and a larger effect of age on speech recognition for complex listening tasks.
Attributing speech recognition declines to aging is simply a way of saying that the deficit goes beyond cochlear hearing loss. Much of the research data has shown that perception and cognition co-vary with age, making it very difficult to distinguish among the peripheral, central auditory, and cognitive deficits that may be occurring with aging (Schneider 2001). Several explanations are used for this covariance: perceptual decline may cause cognitive decline (either through long-term sensory deprivation or short-term information degradation); cognitive decline may cause perceptual decline (cognitive load on perception theory); or there may be an independent third cause for both declines (common-cause theory) (Schneider and Pichora-Fuller 2000).

The interaction between peripheral and central processing also may have some positive aspects for speech recognition. Schneider and Pichora-Fuller (2000) note that declines in bottom-up processing are offset by attention to the broader context, implying that there are ways to compensate for peripheral damage or a degraded input signal. If the sensory input is degraded by hearing loss, the listener may still be able to interpret the message by using higher-level resources. The compensation by other resources adds to the difficulty of isolating the contribution of peripheral damage to speech recognition. For example, speech in one listening condition may be recognized mainly via bottom-up mechanisms, while speech in a degraded listening condition may force the listener to use general knowledge and context (higher-level mechanisms) to recognize the signal. The behavioral scores in terms of percent correct may be the same for both of these conditions, but from the person’s response, there is no way of separating how much of the recognition was due to peripheral contributions and how much was due to central or cognitive contributions.

Let us look at the acoustic and linguistic characteristics of some of the commonly studied complex listening tasks and their effects on speech recognition.

Background noise

Background noise is difficult to summarize because it can constitute any unwanted sound in a listening situation. The properties that may vary include spectral content, modulation depth and frequency, duration, meaningfulness, overall level, and level relative to the signal of interest. A common property of many noise types, however, is masking of low-level speech segments (Nabelek 1993).

For steady state noise, speech recognition is reduced predictably by the reduction in audibility of the speech signal (Dubno et al. 2002; Dubno et al. 1997; Tun and Wingfield 1999). It functions in the same way that peripheral hearing loss does to reduce audibility. This relationship does not hold true at high levels of steady background noise, as the reduction in recognition is greater than expected from the audibility loss (Humes et al. 1994).

For more complex background noise, such as temporally-modulated noise, multi-talker babble, or tone complexes, speech recognition among older listeners does not have a simple, predictable relationship to changes in audibility. In most studies of young-old listeners, the effect of modulated noise has been found to be mostly peripheral and related directly to audibility changes (Gordon-Salant 1987a, 1987b; Gordon-Salant and Fitzgibbons 1995a, 1995b, 1997; Schneider, Daneman, Murphy and See 2000). There are some exceptions, though, in which even in this young-old group it has been found that the difference between the speech recognition score predicted from audibility and the actual score for the young-old listeners is greater in a temporally-modulated background noise than in steady-state noise (Dubno et al. 2002; Sherbecoe and Studebaker 2003). However, Dubno et al. found that the greater deficit for temporally-modulated background noise was related to the listener’s masked thresholds, also a peripheral effect.

For old-old listeners, while peripheral changes in audibility are still important predictors of speech recognition in modulated noise, other factors have a larger role. For example, verbal and non-verbal cognitive factors (Humes 2002), working memory (Pichora-Fuller, Schneider and Daneman 1995), and changes in temporal processing have been associated with reduced speech recognition in modulated noise. Additionally, a non-specific age factor also has been found to be associated with reduced recognition in complex noise. The nature of this factor is such that, once hearing loss has either been controlled or accounted for statistically, a significant portion of the variance in scores can be accounted for by the listener’s age (Frisina and Frisina 1997; Helfer and Huntley 1991; Nabelek 1988).

Meaningful background noise is distinguished from other modulated noise because it likely engages higher-level processes. Noises included in this category are backgrounds of one or two talkers speaking in a language familiar to the listener. Both younger and older listeners benefit when the number of background
talkers is reduced from multiple talkers to a single talker, but older adults do not get as much benefit as younger adults (Tun and Wingfield 1999). The age effect is minimal when the single background talker speaks an unfamiliar language (Tun, O’Kane and Wingfield 2002), indicating that the reduced benefit for older listeners is not simply peripheral masking. This finding is likely related to cognitive factors, which may be either reduced inhibition of the background talker or reduced capacity of processing resources (Tun et al. 2002). The age effect for meaningful background noise is greater for old-old than young-old listeners (Beaman 2005).

Reverberation

Reverberation has two main effects on the acoustic speech signal: first, it causes masking of the signal, thereby reducing audibility, and second, it changes the temporal structure of the speech signal (Helfer and Huntley 1991). Masking of the signal, also known as overlap-masking, occurs when the reflected acoustic signal of one speech segment reduces the audibility of the next speech segment. The amount of masking depends on the relative intensities of the two speech segments and the amount of shared spectral content. Changes to the temporal structure of the signal come from self-masking, when the reflected acoustic signal smears the energy within a phoneme, essentially filling in the low-intensity portions of the signal, thereby reducing the modulation depth (Humes 1993; Nabelek 1993). The change to the temporal structure is greater for higher modulation frequencies than lower ones.

Reverberation has a detrimental effect on recognition even for young-old listeners (Gordon-Salant and Fitzgibbons 1993, 1995a, 1995b), and is greater for old-old listeners (Humes and Christopherson 1991; Nabelek 1988). The effect is particularly evident when reverberation is combined with a second distortion. Individual susceptibility to the negative effects of reverberation is correlated with measures of temporal processing (see Chapter 4 in this proceeding) (Gordon-Salant and Fitzgibbons 1993; Humes and Christopherson 1991).

Other Temporal Manipulations

The finding in many of these studies that reduced speech recognition with increased age was associated with temporal resolution has led to the study of several temporal manipulations of the speech signal. A partial list of these manipulations includes interrupted speech, jittered speech, altered duration of specific segments, and altered amplitude relationships, such as the consonant-vowel ratio (CVR). Reduced recognition by older listeners has been found specifically for interrupted speech (Gordon-Salant and Fitzgibbons 1993). Reduced recognition by listeners of all ages, not just older listeners, have been found for manipulations in which duration or relative amplitude were altered (Gordon-Salant 1986).

One of the most well-studied forms of temporal distortion in older listeners is rapid speech, also known as time-compressed speech. Time-compressed speech is used extensively to investigate the separate contributions of perceptual and cognitive processes to speech recognition in older listeners. In everyday conversation people tend to talk more quickly than the rate used for standard speech testing, so rapid speech is one way of making the lab tests more realistic. For purposes of better experimental control over the variables that may change when a talker adjusts the rate, time-compressed speech is often used to simulate faster talking rates. Time compression of speech is implemented by removing regular intervals from the speech signal and abutting the remaining segments using a weighted overlap of the segments to eliminate clicks and distortions in the signal. This method does not change the pitch of speech and sounds quite natural and intelligible (Arons 1992). As more time compression is applied, representing faster speaking rates, older listeners show greater reductions in speech recognition. This effect is more pronounced when time-compression is combined with processing that distorts the acoustic signal, such as reverberation or complex background noise (Gordon-Salant and Fitzgibbons 1995b, 2004; Tun 1998).

Multiple Distortions

When two or more of these acoustic degradations are combined in a listening situation, particularly when one of the degradations is time compression, the effect for older listeners is almost always greater than would be expected from the results of each acoustic degradation alone (e.g., Gordon-Salant and Fitzgibbons, 1995a; e.g., Gordon-Salant and Fitzgibbons, 1995b). Such a finding indicates that with only a single distortion, the older listener may be taking advantage of the redundancy in the speech signal to compensate for the cues that have been altered by the distortion. The addition of a second distortion removes the possibility of accessing the redundant cues.
Summary

In summary, recognition of distorted or degraded speech by older adults involves factors related to hearing loss and to higher-level processes. Several types of temporal distortions have been shown to have negative consequences on speech recognition, particularly if the listening condition involves multiple distortions, or an impoverished speech signal. There is strong evidence that this effect is not caused only by peripheral hearing loss. The effect increases with age, even within the category of older listeners.

Redundancy of Speech Cues

At the phonetic level, the key to understanding many of these results may lie in knowing that there are multiple acoustic cues available cuing the identity of each segment of the speech signal. Older listeners rely on integrating these acoustic cues for recognition more than younger listeners (Ohde and Abou-Khalil 2001) and they have lower recognition performance when one dimension of the speech signal is restricted (Souza 2000; Souza and Kitch 2001). Older listeners seem to rely on redundancy in the signal, and removing this is detrimental to their speech recognition. For speech in quiet that has not been distorted or manipulated, all acoustic cues are available. As speech is degraded, distorted, or noise is added, some of the acoustic cues to phoneme identity are altered. Younger listeners can use the remaining acoustic cues; older listeners seem to have difficulty using alternate acoustic cues, particularly when cues are in conflict.

Acoustic Cue Manipulation

In acoustic cue manipulation studies, listeners are forced to use minimal acoustic cues to recognize the speech signal (Coughlin, Kewley-Port and Humes 1998; Dorman, Marton, Hannley and Lindholm 1985; Ohde and Abou-Khalil 2001). The stimuli may be created in several possible ways: synthesizing stimuli that vary on only one parameter; putting two acoustic cues in competition to determine the weight that the listener applies to each one; removing portions of the signal, such as the steady-state portion of a vowel while leaving only the dynamic cues (i.e., formant transitions); or providing the listener with a very impoverished acoustic signal. These paradigms have in common that they remove redundancy from the signal to determine which acoustic cues the listeners are weighting the highest for recognition.

Conflicting results have been found from various studies using this approach. However, the general conclusions that can be drawn are that older adults are less able than younger adults to use the dynamic cues of speech as sources of information (Fox, Wall and Gocken 1992), they are less efficient listeners (Ohde and Abou-Khalil 2001), they are less able to use the second formant to identify vowels (Coughlin et al. 1998), and they need to integrate acoustic cues more than younger listeners (Ohde and Abou-Khalil 2001). This last finding means that older listeners need more redundancy in the signal, and in particular, do not perform well when the acoustic cues are in conflict.

Figures 2 and 3 provide an illustration of acoustic cue trade-offs and integration of acoustic cues. Figure 2 shows the waveform and spectrogram for the word “sty”. The two main acoustic cues that signal the presence of the stop consonant /t/ are the silent gap and the release burst. The relevant portions of the acoustic signal have been outlined. In acoustic cue studies, the length of the gap and the amplitude of the release burst would be manipulated to see which cue is most important for speech recognition. It turns out that young listeners with normal hearing can use either cue to detect the presence of the stop consonant /t/ are the silent gap and the release burst. The relevant portions of the acoustic signal have been outlined. In acoustic cue studies, the length of the gap and the amplitude of the release burst would be manipulated to see which cue is most important for speech recognition. It turns out that young listeners with normal hearing can use either cue to detect the presence of the stop consonant, but they benefit from having both; that is, they can detect a stop consonant using only the silent gap, but if the release burst is also present, they can detect the stop consonant with a shorter gap. Old listeners, on the other hand, need both
recognition performance is high and equivalent as long as either the release burst is present or the gap is sufficiently long. By contrast, likely results for old listeners are plotted in the lower panel, indicating that the only listening condition in which recognition would be high for these listeners is when both acoustic cues are available: the long silent gap and the presence of the release burst.

Using Time-Compression to Examine Acoustic Redundancy

As reviewed above, when speech is time-compressed, recognition is reduced, seemingly related to a reduction in available processing time. However, when time-compressed speech is restored to its original length with pauses, older listeners’ performance improves but does not return to its original level (Wingfield and Ducharme 1999). Gordon-Salant and Fitzgibbons (2001) further examined this effect by applying time compression to either an entire sentence or specific parts of the signal. The parts of the signal time-compressed were either consonants only, vowels only, or pauses only. If processing time was the only determinant of the age effect for rapid speech, then the reduction in recognition should be the same for time compression of consonants, vowels, and pauses, provided that the amount of time removed was the same. What Gordon-Salant and Fitzgibbons found was that, although there was an age effect for selective time compression of vowels and pauses, there was a greater age effect for selective time compression of consonants, showing that reduced processing time was an important element of the age effect, but also that the older listeners had greater difficulty processing the rapid acoustic cues or making use of the impoverished acoustic signal.

Summary of Speech Recognition Data

This very brief review of speech recognition by younger and older listeners has shown that, with some individual exceptions, there are no differences in speech recognition performance based on age when speech is presented in quiet and is not temporally distorted. However, age differences begin to be apparent when the listening situation becomes more complex; for example, when speech is distorted or presented in a background of temporally modulated noise. This pattern of findings is consistent with findings that older listeners need access to multiple redundant cues in the speech signal to facilitate recognition. Given these

![Figure 3](image-url)  
**Figure 3.** Hypothetical data illustrating a common finding of a difference between young and old listeners in their use of a trade-off in acoustic cues for speech recognition. When acoustic cues are isolated, young listeners can use either the silent gap or the release burst to detect the presence of a stop consonant. Old listeners need both cues to be congruent to detect the presence of a stop consonant.
results, what are the implications for older listeners using hearing instruments that process the speech signal?

**Acoustic Effects of Hearing Instrument Processing**

**Temporal Changes**

Compression in hearing instruments will be examined in order to understand the consequences of signal processing on older listeners’ use of hearing instruments. Compression has its greatest negative effect upon the temporal envelope of speech, an acoustic dimension that carries important speech information (Van Tasell 1993). Evidence for the importance of the temporal envelope comes from experiments in which a noise carrier is modulated by the speech envelope and listeners are able to reach a good level of performance, suggesting that the envelope provides cues to speech identification (e.g., Shannon, Zeng and Wygonski 1992; Shannon, Zeng, Kamath, Wygonski and Ekelid 1995; Turner, Souza and Forget 1995; Van Tasell, Greenfield, Logemann and Nelson 1992; Van Tasell, Soli, Kirby and Widin 1987). As hearing loss increases, frequency resolution becomes poorer, resulting in a greater reliance on the temporal envelope (e.g., Faulkner, Ball, Rosen, Moore and Fourcin 1992).

The specific information carried by the temporal envelope includes segmental cues to voicing, manner of articulation, and prosodic cues for tempo, rhythm, and syllabic for understanding speech in quiet (Rosen 1989). In the processing of speech in noisy environments, the temporal envelope may have a role in sound source segregation (Crouzet and Ainsworth 2001a, 2001b). When the temporal envelope is smeared, recognition in noise gets worse. This effect is even greater when the background is a modulated noise.

As the time constants of compression (i.e., attack time and release time) are made shorter, the effect on the temporal envelope increases. Specifically, shorter time constants lead to a smoothing of the envelope, such that any level differences that the listener may use to help understand the speech signal are no longer available (Van Tasell 1993). Multiple studies show that fast-acting compression affects the temporal cues in speech (e.g., Boothroyd, Springer, Smith and Schulman 1988; Dreschler 1989; Festen, van Dijkhuizen and Plomp 1990; Plomp 1988; Verschuure et al. 1996). Figure 4 shows the intensity contour (or temporal envelope) of the nonsense syllable /ip/ before and after compression processing. Notice that the shape of the temporal envelope is altered by the processing: differences in level between the vowel and consonant are reduced and the amplitude of the release burst is enhanced with processing.

When the inherent redundancy in the speech signal is reduced by removing most of the spectral information, leaving only temporal cues, then fast-acting wide-dynamic range compression (WDRC) has been shown to reduce recognition for the speech material (Souza 2000;...
Souza and Kitch 2001; Souza and Turner 1996, 1998; Van Tasell and Trine 1996). Fast-acting compression has also been shown to reduce speech quality ratings, even in the absence of changes to objective speech recognition measures (van Buuren, Festen, and Houtgast, 1999).

**Bringing it all Together: A Study of Acoustic Effects of Hearing Instrument Processing, Redundancy, and Specific Speech Errors for Older Listeners**

**Method**

Twenty-five listeners, age 62 to 88 years with moderately to moderately-severely sloping sensorineural hearing loss, participated in the study. The listeners heard and repeated sentences from the Speech Perception in Noise (SPIN; Bilger, Nuetzel, Rabinowitz and Rzeczkowski 1984) test. The sentences have correct syntactic structure but no semantic information to cue the listener to word identity.

Two types of processing were applied to the sentences heard by the listeners: removal of redundant cues and compression processing. Figure 5 shows one example of a sentence processed to remove some of the acoustic cues from the signal. The upper waveform shows the original, unprocessed signal with full redundancy of acoustic cues available. To reduce redundancy, alternate pitch periods were removed from the recording, with the remaining segments as shown in the middle waveform of the figure. Removing alternate pitch periods reduces the length of the stimulus, so time was restored to the sentence by adding in gaps at natural pauses and reduplicating pitch periods of steady-state portions of words. The result of the time restoration procedure is shown on the lower waveform of figure 5. The time restored stimulus is the same length as the original stimulus but with reduced redundancy in the acoustic signal.

The second type of processing was completed using a two-channel compression hearing instrument simulator. Processing parameters were set to create a range of distortions to the intensity contour of the speech signal. Overall, four distortion levels were tested.

The purpose of the two processing types was to determine whether these older listeners could compensate for the distortion produced by hearing instrument processing by using alternate acoustic cues in the speech signal. It was hypothesized that speech recognition would remain high for the full redundancy condition even at high levels of distortion, but that speech recognition would be detrimentally affected by increasing levels of distortion when acoustic redundancy was limited.
Results and Interpretation

The results of speech recognition for these sentences are shown in figure 6. It can be seen on the figure (and this was confirmed statistically) that the detrimental effect of distortion was much greater for the reduced redundancy condition than the condition in which full acoustic cues were available. Thus it is likely that these older listeners were using alternate acoustic cues in the signal to compensate for distortion of the intensity contour in the speech signals when these cues were available. When the acoustic cues were limited by removing alternate pitch periods, these listeners were not able to compensate as well for the distortion.

Figures 7 and 8 show examples of tokens before and after hearing instrument processing. Both tokens have stop consonants and the second token, “ditch” contains an affricate. The main characteristics to notice are that in the processed condition the release bursts are enhanced and the silent gap portion of the stops now contains some noise. The behavioral consequences of these acoustic changes can be seen in table 1.

Table 1 gives the percentage of errors for stops and affricates when full acoustic cues are available and when redundancy is reduced. For the full acoustic redundancy condition, the effect of distortion was very small, but for the reduced redundancy condition, percent correct recognition decreased as distortion increased. The final column of table 1 lists the most common types of errors that were made in the high distortion/reduced redundancy condition. That is, what did the listeners perceive when the available acoustic cues were distorted and redundant cues were not available to help them compensate? This listening condition is similar to the conditions described earlier, in which gaps and release bursts were traded off.

For voiceless stops, the most common error was substitution of a voiced stop. This may have been due to the presence of increased noise during the stop closure,
Figure 8. Time waveform and spectrogram of the word “ditch” unprocessed (left panel) and processed (right panel). See text for further explanation.

Table 1. Specific errors for stops and affricates listed as a function of available redundancy (full cues or reduced redundancy) and amount of acoustic distortion (low or high distortion). Errors for word initial and word final phonemes are combined.

<table>
<thead>
<tr>
<th></th>
<th>Distortion Level</th>
<th>Full Cues Percent</th>
<th>Reduced Redundancy Percent</th>
<th>Specific Errors in the Reduced Redundancy Condition</th>
</tr>
</thead>
</table>
| **Voiceless Stops /p t k/** | Low Distortion   | 90.4              | 79.5                      | 1) Errors of voicing  
2) Errors of place of articulation  
3) Deletion of the phoneme  
4) Errors of manner – substituted a fricative |
|                         | High Distortion  | 86.6              | 47.8                      |                                                      |
| **Voiced Stops /b d g/** | Low Distortion   | 93.3              | 73.6                      | 1) Errors of voicing  
2) Deletion of the phoneme |
|                         | High Distortion  | 90.8              | 57.8                      |                                                      |
| **Affricates /tS dZ/**  | Low Distortion   | 87.2              | 90.0                      | 1) Errors of manner – substituted a fricative  
2) Deletion of the phoneme |
|                         | High Distortion  | 80.0              | 60.0                      |                                                      |
which listeners interpreted as the presence of voicing. Errors of place of articulation were common. Place of articulation is cued in part by the amplitude of the release burst, which we know was altered by processing. Another common error was deletion of the stop altogether – with reduced or altered cues, the stop was no longer detected. And finally, several listeners responded with a fricative instead of a stop; this may have been due to the much increased amplitude of the release burst, which was now perceived as frication noise.

For voiced stops, the most common errors were substitution of a voiceless stop or simply deletion of the phoneme. Alterations of the acoustic cues may have reduced cues to voice onset time, leading to the perception of a voiceless stop, or cues were reduced to the point at which a stop was no longer perceived.

For affricates, which are essentially a blend of a stop consonant followed by a fricative with a very fast rise time, the most common error was substitution of a fricative. Alteration of the cues reduced access to the timing cue of the fast rise time that signaled the affricate rather than the fricative. Another common error was simple deletion of the entire phoneme, indicating that the listeners did not have enough undistorted acoustic information remaining to provide a reasonable estimate of the phoneme’s identity.

The pattern of results obtained here is consistent with what we might expect based on the earlier review of speech recognition. That is, alteration of one or more of the acoustic cues important for recognition of stop consonants affects listeners’ perception of the speech stimulus. Enhancement of the release burst led to errors of place of articulation and even manner, since several listeners responded with a fricative instead of a stop. The errors and effects of distortion were greater when redundancy was removed from the speech stimulus, confirming that older listeners are relying on redundant acoustic cues in the speech signal.

Conclusions and Clinical Implications

Taking all of the above discussion together, what does this mean for clinical practice in audiology? First, many older listeners have difficulty with speech recognition when the listening situation is complex; for example, when background noise is fluctuating or when the speech signal is distorted, as with reverberation. Clinically this has important implications for counseling clients about their own reasonable expectations in difficult listening situations, and may lead the clinician to recommend other hearing assistance technology that may improve the signal to noise ratio.

It is possible to alter the amplitude cues of speech with WDRC processing to some extent without affecting speech recognition. WDRC provides audibility, particularly for low-level speech sounds. Increases in audibility are positive for speech recognition, as more acoustic cues (and therefore greater redundancy) are available to listeners. When redundant acoustic cues are available in the speech signal, older listeners are quite tolerant of alterations to the temporal envelope. As redundancy is removed, older listeners have reduced speech recognition when the temporal envelope (or intensity contour) of speech is altered. The specific errors made are predictable from what is known of speech recognition in older listeners, particularly their need to integrate multiple acoustic cues to the identity of a speech segment.

The example of WDRC was used in the study reported here, but the same principle applies to other types of processing in hearing instruments. I encourage clinicians to think about the acoustic effects on speech of new processing schemes as they become available in hearing instruments.

Acknowledgements

Portions of the research presented in this paper were funded by the Canadian Institutes of Health Research. The author wishes to thank Marc Caldwell for his help in stimulus preparation and data collection, and Jennifer Parker and Alison Beers for their help with the speech error analysis and acoustic analysis of the stimuli.

References


