27

Lynne E. Bernstein & Edward T. Auer, Jr.

Speech Perception and Spoken Word Recognition

Speech is an important mode of communication for many people with hearing losses, even with losses at severe (60–89 dB HL) or profound (>90 dB HL bilaterally) levels. Individuals with hearing losses of these magnitudes occupy positions on a continuum between relying exclusively on spoken language and relying exclusively on manual language. Speech perception can depend totally on heard speech at one extreme and on seen speech (lipreading/speechreading) at the other.[1] In addition, communication conditions can determine where on the continuum an individual is at any particular time. For example, students at Gallaudet University who relied on manual language in their classrooms and elsewhere on campus reported reliance on spoken language for communicating with their hearing friends, families, and the public (Bernstein, Demorest, & Tucker, 1998).

This chapter focuses on spoken communication by adults with severe or profound hearing loss, although it includes relevant discussion of results from studies involving participants with mild to moderate hearing losses or with normal hearing. The chapter describes several fundamental issues in speech perception and spoken word recognition and reviews what is known about these issues in relation to perceivers with severe-to-profound hearing losses.

Speech Perception

When talkers produce speech, their articulatory gestures typically produce acoustic and optical signals that are available to perceivers. The auditory and visual perceptual systems must categorize the linguistically relevant speech information in the speech signals. The physical forms of speech have a hierarchical structure. The segmental consonants and vowels comprise subsegmental features. Those features can be described in articulatory terms such as place of articulation (e.g., bilabial, dental, alveolar), manner of articulation (e.g., stop, liquid, vocalic, nasal), and voicing (voiced, unvoiced) (Catford, 1977).[2] The speech segments are used in language combinatorially to form morphemes (minimal units of linguistic analysis such as "un," "reason," "able" in "unreasonable"), which in turn combine to form words. Language differs from other animal communication systems in its generativity, not only to produce infinitely many different sentences out of a set of words but also to gen-

erate new words by combining the finite set of segmental consonants and vowels within a particular language.

That consonants and vowels are structurally key to the generation of word forms has also suggested that they are key to the perception of words. However, discovering how perceivers recognize the consonant and vowel segments in the speech signals produced by talkers has not proved straightforward and has not yet been fully accomplished (e.g., Fowler, 1986; Liberman & Whalen, 2000; Nearey, 1997). The reason for this difficulty is that the speech segments are not produced like beads on a string, and so do not appear as beads on a string in the acoustic signal (Liberman, 1982). The speech articulators—the lips, tongue, velum, and larynx—produce speech gestures in a coordinated and overlapping manner that results in overlapping information. The speech production gestures change the overall shape of the vocal tract tube, and those shapes are directly responsible for the resonances (formants/concentrations of energy) of the speech signal (Stevens, 1998). However, different vocal tract shapes can produce signals that are perceived as the same segment, further complicating matters.

Numerous experiments have been conducted using synthesized, filtered, and edited speech waveforms to isolate the parts of the speech signal that are critical to the perception of speech. Although it is not yet completely known how auditory perceptual processes analyze acoustic speech signals, it is known that listeners are remarkably capable of perceiving the linguistically relevant information in even highly degraded signals (e.g., Remez, 1994). The questions of importance here are what auditory information can be obtained by individuals with severe or profound hearing loss and how speech perception is affected by individual hearing loss configurations. Work on this problem began with examining how speech perception with normal hearing is affected by various manipulations such as filtering. For example, Miller and Nicely (1955) showed that perception of place of articulation (e.g., /b/ versus /d/ versus /g/) information depends greatly on the frequencies above 1000 Hz, but voicing (e.g., /b/ versus /p/) is well preserved with only frequencies below 1000 Hz. The manner feature involves the entire range of speech frequencies and appears to be less sensitive to losses in the higher frequencies.

Auditory-only Speech Perception of Listeners with Impaired Hearing

As level of hearing loss increases, access to auditory speech signals decreases. At severe or profound levels of hearing loss, hearing aids can help overcome problems with audibility of speech sounds for some individuals, particularly when listening conditions are clear. Amplification systems are designed to restore audibility by boosting intensity in regions of the spectrum affected by the loss. Unfortunately, when hearing loss is severe or profound, simply increasing the amplitude of the signal does not always restore the listener's access to the information in the speech signal: At those levels of hearing loss, the speech information that can be perceived auditorily is typically highly degraded due to limitations imposed by the listener's auditory system. For example, high sound-pressure levels required to amplify speech adequately to compensate for severe or profound levels result in additional distortion, apparently equivalent to the distortion experienced by hearing people under equivalent signal presentation conditions (Ching, Dillon, & Byrne 1998). However, it is difficult to generalize across individuals. Results vary, and many different factors may be involved in how well a hearing aid ameliorates the effects of the hearing loss. These factors include the specific type of hearing loss (e.g., the specific frequencies and the magnitude of the loss for those frequencies), and, quite likely, factors involving central brain processing of the auditory information, including word knowledge and experience listening to the talker.

Specific speech features are affected at different levels of hearing loss. Boothroyd (1984) conducted a study of 120 middle- and upper-school children in the Clarke School for the Deaf in Northampton, Massachusetts. The children's hearing losses, measured in terms of pure-tone averages in decibels of hearing level (dB HL) ranged between 55 and 123 dB. The children were tested using a four-alternative, forced-choice procedure for several speech segment contrasts. The results showed that as the hearing losses increased, specific types of speech contrasts became inaudible, but information continued to be available even with profound losses. After correcting for chance, the point at which scores fell to 50% was 75 dB HL for consonant place, 85 dB HL for initial consonant voicing, 90 dB HL for initial consonant continuance, 100

dB HL for vowel place (front-back), and 115 dB HL for vowel height. Boothroyd thought these might be conservative estimates of the children's listening abilities, given that their hearing aids might not have been optimized for their listening abilities.

Ching et al. (1998) reported on a study of listeners with normal hearing and listeners with hearing losses across the range from mild to profound. They presented sentence materials for listening under a range of filter and intensity level conditions. Listeners were asked to repeat each sentence after its presentation. Under the more favorable listening conditions for the listeners with severe or profound losses, performance scores covered the range from no words correct to highly accurate (approximately 80–90% correct). That is, having a severe or profound hearing loss was not highly predictive of the speech identification score, and some listeners were quite accurate in repeating the sentences. In general, the majority of the listeners, including listeners whose hearing losses were in the range of 90–100 dB HL (i.e., with profound losses), benefited from amplification of stimuli for the frequencies below approximately 2800 Hz. (Telephones present frequencies in a range only up to approximately 3200 Hz, suggesting that perceiving frequencies up to 2800 could be very useful.)

Turner and Brus (2001) were interested in the finding that when hearing loss is greater than 40–80 dB HL for the higher frequencies of speech, very little benefit is achieved by increasing the amplification of those higher frequencies, and, in some cases, the amplification actually results in lower performance. However, amplification of lower frequency regions does seem to provide benefit. They hypothesized that there might be an interaction between effects due to the frequency regions for which hearing loss occurred and the types of speech information the listeners were able to perceive, depending on amplification characteristics. Listeners who had hearing losses from mild to severe were asked to identify consonant-vowel and vowel-consonant nonsense syllables that were low-pass filtered at the cutoff frequencies of 560, 700, 900, 1120, 1400, 2250, and 2800 Hz. That is, only the frequencies below the cutoff were in the stimuli.

A main question for Turner and Brus (2001) was whether amplification of the lower frequencies of speech was helpful regardless of the level of hearing loss; affirmative findings were obtained across listeners and filter conditions. Turner and Brus also analyzed their data to determine how the speech features of manner, voicing, and place were independently affected by the filtering conditions and the degree of hearing loss. The manner feature refers to the distinction between consonants that are stops (e.g., /b, d, g/) versus fricatives (e.g., /f, s, z/), versus affricates (e.g., /j, č/), versus liquids (e.g., /l, r/). For this feature, performance generally improved as the filter cutoff allowed more frequencies into the stimuli. The voicing feature refers to the distinction between voiced (e.g., /b, d, g/) and voiceless (e.g., /p, t, k/) consonants. This feature was transmitted well to all the listeners, even when the low-pass filter cutoff was at its lowest levels, and even for the listeners with the more severe losses. That is, the voicing cue is robust to extreme limitations in the low frequency range of audible speech. The place feature refers to the position in the vocal tract where the consonant occlusion is formed (e.g., /b/ is formed by closure of the lips and /k/ is formed by closure of the back portion of the tongue against the velum). This feature was most sensitive to addition of higher frequencies and was most sensitive to the degree of hearing loss. Listeners with the more severe losses were unable to benefit much as additional higher frequencies were allowed into the stimulus.

In general, Turner and Brus (2001) confirmed the Ching et al. (1998) findings, suggesting that listeners with severe or profound hearing loss benefit most from amplification of the lower frequencies of speech. Nevertheless, in comparisons with hearing listeners, amplification for those with severe or profound hearing losses does not restore speech perception accuracy to normal levels.

Lipreading

As the level of hearing loss increases, and/or in environmental noise increase, people with severe or profound hearing losses typically must rely on being able to see visual speech information to augment or substitute for auditory speech information. The literature on lipreading does not necessarily encourage the view that visual information is a good substitute for auditory information. Estimates of the upper extremes for the accuracy of lipreading words in sentences have been as low as 10–30% words correct (Rönnberg, 1995; Rönnberg, Samuelsson, & Lyxell, 1998). Estimates of the ability

to perceive consonants and vowels via lipreading alone have varied across studies and the particular stimuli used. Such studies typically involve presentation of a set of nonsense syllables with varied consonants or varied vowels and a forced-choice identification procedure. In general, consonant identification is reported to be less than 50% correct (e.g., Owens & Blazek, 1985), and vowel identification is reported to be somewhat greater than 50% correct (e.g., Montgomery & Jackson, 1983).

Several authors have asserted that the necessity to rely on visible speech due to hearing loss does not result in enhanced lipreading performance (e.g., Summerfield, 1991), and that lipreading in hearing people is actually better than in deaf people due to auditory experience in the former (Mogford, 1987). Furthermore, several authors assert that lipreaders can only perceive visemes (e.g., Fisher, 1968; Massaro, 1987, 1998). That is, the consonant categories of speech are so highly ambiguous to lipreaders that they can only distinguish broadly among groups of consonants, those broad groups referred to as visemes. Finally, some estimates of how words appear to lipreaders have suggested that approximately 50% of words in English appear to be ambiguous with other words (Berger, 1972; Nitchie, 1916).

To investigate some of these generalizations, Bernstein, Demorest, and Tucker (2000) conducted a study of lipreading in 96 hearing students at the University of Maryland and in 72 college students at Gallaudet University with 60 dB HL or greater bilateral hearing losses. All of the Gallaudet students reported English as their native language and the language of their family, and they had been educated in a mainstream and/or oral program for 8 or more years. Seventy-one percent of the students had profound hearing losses bilaterally. Sixty-two percent had hearing losses by age 6 months. The participants were asked to lipread nonsense syllables in a forced-choice procedure and isolated words and sentences in an open set procedure. The stimuli were spoken by two different talkers who were recorded on laser video disc.

Results of the study revealed a somewhat different picture of lipreading from that of previous studies. Across all the performance measures in this study, deaf college students were significantly more accurate than were the hearing adults. Approximately 65–75% of the deaf students outperformed 75% of the hearing students. The entire upper quartile of deaf students' scores was typically above the upper quartile of hearing students' scores. For example, one sentence set produced upper quartile scores of percent correct words ranging between 44 and 69% for the hearing students and ranging between 73 and 88% for the deaf students. When the results were investigated in terms of the perceptual errors that were made during lipreading of sentences, the deaf students were far more systematic than the hearing students: when deaf students erred perceptually, they were nevertheless closer to being correct than were the hearing students. When the nonsense syllable data were analyzed in terms of the subsegmental (subphonemic) features perceived, the results showed that the deaf students perceived more of the features than did the hearing students. Finally, among those deaf students with the highest performance were ones with profound, congenital hearing losses, suggesting that visual speech perception had been the basis for their acquisition of knowledge of spoken language, and that reliance on visible speech can result in enhanced perceptual ability.

Bernstein, Demorest, et al. (1998) investigated possible correlations between lipreading performance levels in the Bernstein et al. (2000) study and other factors that might affect or be related to visual speech perception. They examined more than 29 variables in relationship to the deaf students' identification scores on nonsense syllables, isolated words, and isolated sentences. The broad categories of factors that they investigated included audiological variables, parents' educational levels, home communication practices, public communication practices, self-assessed ability to understand via speech, self-assessed ability to be understood via speech, and scores on the Gallaudet University English Placement Test. The parents' educational levels were found not to be correlated with lipreading scores. Neither were most of the audiological variables, such as when the hearing loss occurred, when it was discovered, or level of hearing loss.

Important variables related to lipreading scores included (1) frequency of hearing aid use, which was generally positively correlated with speech scores, such that the more frequently the hearing aid was used the more accurate the student's lipreading (r ranged from .350 to .384);[3] (2) communication at home with speech, which was correlated with better lipreading scores (r ranged from .406 to .611); (3) self-assessed ability to be under-

stood via speech in communication with the general public (r ranged from .214 to .434); and (4) the reading subtest of the English Placement Test (r ranged from .257 to .399).

Regression analyses were used to investigate the best predictors of lipreading scores among the variables that produced significant correlations. Only three factors survived the analysis as the significant predictors for scores on words and sentences: self-assessed ability to understand the general public, communication at home with speech, and the English Placement Test score. In fact, the multiple R values obtained from the analysis were quite high, ranging from .730 to .774 for scores on lipreading words and sentences. That is, more than 50% of the variance in the scores was accounted for by the three best factors. To summarize, lipreading ability was highly related to experience communicating successfully via speech and was also related to the ability to read.

Spoken Word Recognition

The focus on perception of the segmental consonants and vowels in the speech perception literature might leave the reader with the impression that perception of speech terminates in recognition of the speech segments. Indeed, some researchers theorize that perception of spoken language involves perceptual evaluation of subsegmental units to categorize the consonant and vowel segments at an abstract level (e.g., Massaro, 1998). Recognition of words would then depend on assembling the abstract segmental categories and matching them to the segmental patterns of words in long-term memory. According to this view, perception terminates at the level of recognizing segments. However, research on spoken word recognition suggests that perception extends to the level of lexical processing.

Abundant evidence has been obtained showing that the speed and ease of recognizing a spoken word is a function of both its phonetic/stimulus properties (e.g., segmental intelligibility) and its lexical properties (e.g., "neighborhood density," the number of words an individual knows that are perceptually similar to a stimulus word, and "word frequency," an estimate of the quantity of experience an individual has with a particular word) (Lahiri & Marslen-Wilson, 1991; Luce, 1986; Luce, & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990; Marslen-Wilson, 1992; McClelland & Elman, 1986; Norris, 1994).

"Segmental intelligibility" refers to how easily the segments (consonants and vowels) are identified by the perceiver. This is the factor that segmental studies of speech perception are concerned with. Word recognition tends to be more difficult when segmental intelligibility is low and more difficult for words that are perceptually similar to many other words (see below). This factor shows that perception does not terminate at the level of abstract segmental categories. If perception did terminate at that level, it would be difficult to explain stimulus-based word similarity effects. Word recognition tends to be easier for words that are or have been experienced frequently. This factor might be related to perception or it might be related to higher level decision-making processes. All of these factors have potential to be affected by a hearing loss.

General Theoretical Perspective

Theories in the field of spoken word recognition attempt to account for all the factors defined above within a framework that posits perceptual (bottom-up) activation of multiple word candidates. Activation is a theoretical construct in perception research but is thought to be directly related to activation of relevant neural structures in the brain. The level of a word's bottom-up activation is a function of the similarity between the word's perceptual representation and that of candidate word forms stored in long-term memory (e.g., Luce, 1986; Luce, Goldinger, Auer, & Vitevitch, 2000; Luce & Pisoni, 1998; Marslen-Wilson, 1987, 1990; McClelland & Elman, 1986; Norris, 1994). Once active, candidate word forms compete for recognition in memory (Luce, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1992; McClelland & Elman, 1986; Norris, 1994). In addition to bottom-up stimulus information, recognition of a word is influenced by the amount and perhaps the type of previous experience an individual has had with that word (Goldinger, 1998; Howes, 1957). It is important to emphasize here that the long-term memory representations of stimulus word forms are hypothesized to be similar to the perceptual information and therefore different from memory representations for other types of language input (e.g., fingerspelling), as well as different from abstract

knowledge about words (e.g., semantics; McEvoy, Marschark, & Nelson, 1999).

An implication of the view that the perceptual word information is used to discriminate among words in the mental dictionary (lexicon) is that successful word recognition can occur even when the speech signal is degraded. This is because recognition can occur even when the speech signal contains only sufficient information to discriminate among the word forms stored in the mental lexicon. For example, an individual with hearing loss may distinguish the consonants /p/, /t/, and /k/ from the other segments in English but might not distinguish within this set. For this individual, the word "parse" could still be recognized because "tarse" and "karse" do not occur as words in English. That is, words are recognized within the context of perceptually similar words, and therefore intelligibility is a function of both segmental intelligibility as well as the distribution of word forms in the perceiver's mental lexicon.

Visually Identifying Words with Reduced Speech Information

One fundamental question is what effect reduced speech information, such as the information available to the lipreader, has on the patterns of stimulus words that are stored in the mental lexicon. Nitchie (1916) and Berger (1972) investigated the relationship between reduced segmental intelligibility and the distribution of word forms for individuals with profound hearing losses who relied primarily on visible speech for oral communication. They argued that as a result of low consonant and vowel accuracy during lipreading, approximately 50% of words in English that sound different lose their distinctiveness (become homophenous/ambiguous with other words).

Auer and Bernstein (1997) developed computational methods to study this issue for lipreading and any other degraded perceptual conditions for speech. They wondered to what extent words lost their distinctive information when lipread—that is, how loss of distinction would interact with the word patterns in the mental dictionary. For example, even though /b/, /m/, and /p/ are similar to the lipreader, English has only the word, "bought," and not the words "mought" and "pought." So "bought" remains a distinct pattern as a word in English, even for the lipreader.

Specifically, the method incorporates rules to transcribe words so that only the segmental distinctions that are estimated to be perceivable are represented in the transcriptions. The rules comprise mappings for which one symbol is used to represent all the phonemes that are indistinct to the lipreader.[4] Then the mappings are applied to a computer-readable lexicon. For example, /b/ and /p/ are difficult to distinguish for a lipreader. So, words like "bat" and "pat" would be transcribed to be identical using a new common symbol like B (e.g., "bat" is transcribed as BAT and "pat" is transcribed as BAT). Then the transcribed words are sorted so that words rendered identical (no longer notationally distinct) are grouped together. The computer-readable lexicon used in these modeling studies was the PhLex lexicon. PhLex is a computer-readable phonemically transcribed lexicon with 35,000 words. The words include the 19,052 most frequent words in the Brown corpus (a compilation of approximately 1 million words in texts; Kucera & Francis, 1967).

Auer and Bernstein (1997) showed that when all the English phonemes were grouped according to the confusions made by average hearing lipreaders (i.e., the groups /u, ʊ, ər/, /o, aʊ/, /ɪ, i, e, ɛ, æ/, /ɔ, ɪ/, /ɔ, aɪ, ə, ɑ, ʌ, j/, /b, p, m/, /f, v/, /l, n, k, ŋ, g, h/, /d, t, s, z/, /w, r/, /ð, θ/, and /ʃ, tʃ, ʒ, dʒ/), 54% of words were still distinct across the entire PhLex lexicon. With 19 phoneme groups, approximately 75% of words were distinct, approximating an excellent deaf lipreader. In other words, small perceptual enhancements will lead to large increases in lipreading accuracy.

In addition to computational investigations of the lexicon, lexical modeling provides a method for generating explicit predictions about word identification accuracy. For example, Mattys, Bernstein, and Auer (2002) tested whether the number of words that a particular word might be confused with affects lipreading accuracy. Deaf and hearing individuals who were screened for above-average lipreading identified visual spoken words presented in isolation. Results showed that identification accuracy across deaf versus hearing participant groups was not different. The prediction that words would be more difficult, if there were more words with which they might be confused, was born out: Word identification accuracy decreased as a negative function of increased number of words estimated to be similar to the lipreader. Also, words

with higher frequency of occurrence were easier to lipread.

In another related study, Auer (2002) applied the neighborhood activation model (NAM) of auditory spoken word recognition (Luce, 1986; Luce, & Pisoni, 1998) to the prediction of visual spoken word identification. The NAM can be used to obtain a value that predicts the relative intelligibility of specific words. High values are associated with more intelligible words. Deaf and hearing participants identified visual spoken words presented in isolation. The pattern of results was similar across the two participant groups. The obtained results were significantly correlated with the predicted intelligibility scores (hearing: $r = .44$; deaf: $r = .48$). Words with many neighbors were more difficult to identify than words with few neighbors. One question that might be asked is whether confusions among words really depends on the physical stimuli as opposed to their abstract linguistic structure. Auer correlated the lipreading results with results predicted on the basis of phoneme confusion patterns from identification of acoustic speech in noise, a condition that produces different patterns of phoneme confusions from those in lipreading. When the auditory confusions replaced the visual confusions in the computational model, the correlations were no longer significant. This result would be difficult to understand if word recognition were based on abstract phoneme patterns and not on the visual speech information.

Auditorily Identifying Words Under Conditions of Hearing Loss

The NAM has also been used to investigate auditory spoken word recognition in older listeners (52–84 years of age) with mild to moderate hearing loss (Dirks, Takayanagi, Moshfegh, Noffsinger, & Fausti, 2001). Words were presented for identification from word lists that varied the factors of neighborhood density (word form similarity), mean neighborhood frequency (frequency of occurrence of words in the neighborhood), and word frequency. All of the factors were significant in the results. Overall, high-frequency words were identified more accurately than low-frequency words. Words in low-density neighborhoods (few similar neighbors) were recognized more frequently than words in high-density neighborhoods. Words in neighborhoods of words that were generally low in frequency were recognized more accurately than words in neighborhoods of words that were generally high in frequency. The pattern of results was overall essentially similar to results with a different group of listeners with normal hearing. However, the difference between best and worst conditions for listeners with hearing losses (20 percentage points) was greater than for listeners with normal hearing (15 percentage points). This difference among listeners suggests that lexical factors may become more important as listening becomes more difficult. Although the participants in this study had mild to moderate hearing losses, the study suggests that the processes of spoken word recognition are substantially similar across listeners.

In a related study, characteristics of the listeners included hearing loss versus normal hearing and native versus non-native listeners to English (Takayanagi, Dirks, & Moshfegh, in press). Participants were 20 native listeners of English with normal hearing, 20 native listeners with hearing loss, 20 non-native listeners with normal hearing, and 20 non-native listeners with hearing loss. Hearing losses were bilateral and mild to moderate. In this study, there were two groups of words, ones with high word frequency and in low-density neighborhoods (easy words), and ones with low word frequency and in high-density neighborhoods (hard words). Familiarity ratings were obtained on each of the words from each of the participants to statistically control for differences in long-term language experience. In general, there were significant effects obtained for hearing differences and for native language differences: listeners with normal hearing were more accurate than listeners with hearing losses, and native listeners were more accurate than non-native listeners. Easy words were in fact easier than hard words for all of the listeners. However, the difference between native and non-native listeners was greater for the easy words than for the hard words. These results suggest that the neighborhood structure affects both native and non-native listeners, with and without hearing losses. Additional analyses showed that important factors in accounting for the results included the audibility of the words (how loud they had to be to be heard correctly) and also the listener's subjective rating of their familiarity with each of the words.

Estimating Lexical Knowledge

An individual's knowledge of words arises as a function of his or her linguistic experience. Several variables related to lexical experience have been demonstrated to have some impact on the word recognition process, including the age at which words are acquired, the form of the language input (e.g., spoken or printed), and the frequency of experience with specific words (as discussed earlier). Prelingually deaf individuals' linguistic experience varies along all of these dimensions. Impoverishment in the available auditory information typically leads to delayed acquisition of a spoken language, often resulting in reductions in total exposure to spoken language. Prelingually deaf individuals are also likely to use some form of manual communication as their preferred communication mode, and/or as a supplement to lipreading. Several forms of manual communication can fulfill this role, including a form of English-based signing, American Sign Language (ASL), and cued speech (see Leybaert & Alegria, this volume). As a result of variation in these experiential factors, the prelingually deaf population comprises individuals who differ dramatically in the quantity and quality of their perceptual and linguistic experience with spoken words.

In this section, some studies are discussed that focused on lexical knowledge in expert lipreaders. The participants were all individuals who reported English as their native language and as the language of the family, were educated in a mainstream and/or oral program for 8 or more years, and were skilled as lipreaders.

Estimates of the relative quantity of word experience for undergraduates with normal hearing are based on objective word frequency counts based on text corpora (e.g., Kucera & Francis, 1967). However, this approach has its detractors, especially for estimating experience with words that occur infrequently in the language (Gernsbacher, 1984). Furthermore, the approach is clearly insensitive to individual differences that may occur within or between populations of English language users with different lexical experience.

An alternative to using objective counts to estimate word experience is to collect subjective familiarity ratings by having participants rate their familiarity with words presented individually using a labeled scale. Although several sources of knowledge likely contribute to these ratings, general agreement exists that familiarity partly reflects quantity of exposure to individual words. Auer, Bernstein, and Tucker (2000) compared and contrasted familiarity ratings collected from 50 hearing and 50 deaf college students. Judgments were made on a labeled scale from 1 (never seen, heard, or read the word before) to 7 (know the word and confident of its meaning). The within-group item ratings were similar ($r = .90$) for the two participant groups. However, deaf participants consistently judged words to be less familiar than did hearing participants.

Another difference between the groups emerged upon more detailed analysis of the ratings within and across participant groups. Each participant group was split into 5 subgroups of 10 randomly selected participants. Mean item ratings for each subgroup were then correlated with those of the other nine subgroups (four within a participant group and five between). The correlation coefficients were always highest within a participant group. That is, deaf participants used the familiarity scale more like other deaf participants than like hearing participants. The results suggested that despite the global similarity between the two participant groups noted above, the two groups appear to have experienced different ambient language samples. Thus, these results point to the importance of taking into account experiential differences in studies of spoken word recognition.

Another factor in the developmental history of an individual's lexicon is the age at which words are acquired. The age of acquisition (AOA) effect—faster and more accurate recognition and production of earlier acquired words—has been demonstrated in hearing participants using several measures of lexical processing (for a review, see Morrison & Ellis, 1995). Ideally, AOA for words would be based on some objective measure of when specific words were learned. However, AOA is typically estimated by the subjective ratings of adults. These ratings have been shown to have both high reliability among raters and high validity when compared to objective measures of word acquisition (Gilhooly & Gilhooly, 1980).

Auer and Bernstein (2002) investigated the impact of prelingual hearing loss on AOA. In this study, 50 hearing and 50 deaf participants judged AOA for the 175 words in form M of the Peabody Picture Vocabulary Test-Revised (PPVT; Dunn &

Dunn, 1981) using an 11-point scale labeled both with age in years and a schooling level. In addition, the participants rated whether the words were acquired through speech, sign language, or orthography.

The average AOA ratings for stimulus items were highly correlated across participant groups (r = .97) and with the normative order in the PPVT (r = .95 for the deaf group, and r = .95 for the hearing group), suggesting that the groups rated the words as learned in the same order as the PPVT assumes. However, the two groups differed in when (~ 1.5 years difference on average), and how (hearing: 70% speech and 30% orthography; deaf: 38% speech, 45% orthography, 17% sign language) words were judged to have been acquired. Interestingly, a significant correlation (r = .43) was obtained in the deaf participant group between the percent words correct on a lipreading screening test and the percentage of words an individual reported as having been learned through spoken language, with the better lipreaders reporting more words learned through spoken language. Taken together, the results suggested that despite global similarity between the two participant groups, they appear to have learned words at different times and through different language modes.

Bimodal Speech Perception

The preceding sections reveal that individuals with severe or profound hearing losses can potentially obtain substantial speech information from auditory-only or visual-only speech stimuli. That visual speech can substantially enhance perception of auditory speech has been shown with listeners having normal hearing and hearing losses (e.g., Grant, Walden, & Seitz, 1998; Sumby & Pollack, 1954).

Estimates of how audiovisual speech stimuli can improve speech perception have been obtained from children and adults with hearing losses. Lamoré, Huiskamp, van Son, Bosman, and Smoorenburg (1998) studied 32 children with pure-tone average hearing losses in a narrow range around 90 dB HL. They presented the children with consonant-vowel-consonant stimuli and asked them to say and write down exactly what they heard, saw, or heard and saw. Extensive analyses of the results were provided, but of particular interest here were the mean scores for totally correct responses in the auditory-only, visual-only, and audiovisual conditions. When the children were subdivided into groups according to their pure-tone averages, the group with the least hearing losses (mean 85.9 dB HL) scored 80% correct auditory-only, 58% visual-only, and 93% audiovisual. The group with the greatest hearing losses (mean 94.0 dB HL) scored 30% auditory-only, 53% visual only, and 74% audiovisual. The audiovisual combination of speech information was helpful at both levels, but especially for those with the greater hearing loss.

Grant et al. (1998) presented auditory, visual, and audiovisual sentence stimuli to adult listeners from across a range of hearing losses from mild to severe. Overall, sentence scores were audiovisual, 23–94% key words correct, audio only, 5–70% key words correct, and visual only, 0–20% key words correct. Every one of the listeners was able to improve performance when the stimuli were audiovisual. This was true even when the lipreading-only stimuli resulted in 0% correct scores. Benefit from being able to see the talker was calculated for each participant (benefit = (AV − A)/(100 − A); A = audio only, AV = audiovisual). Across individuals, the variation was large in the ability to benefit from the audiovisual combinations of speech information: the mean benefit was 44% with a range from 8.5–83%.

That even highly degraded auditory information can provide substantial benefit in combination with lipreading has also been shown in adult listeners with normal hearing. Breeuwer and Plomp (1984) presented spoken sentences visually in combination with a range of processed auditory signals based on speech. Lipreading scores for the sentences were approximately 18% words correct. One particularly useful auditory signal combined with lipreading was a 500-Hz pure tone whose amplitude changed as a function of the amplitude in the original speech around that frequency. When this signal was combined with lipreading, the mean score for the audiovisual combination was 66% percent words correct. When the same stimulus was then combined with another pure tone at 3160 Hz, also changing in amplitude as a function of the amplitude changes in the original speech around that frequency, performance rose to a mean of 87% words correct.

For neither type of auditory signal alone would there likely have been any words correctly identified. These results demonstrate that being able to hear even extremely limited speech information can be effective, as long as it is combined with visual speech.

Vibrotactile Cues

Under certain conditions, a hearing aid could provide useful vibrotactile information that could combine with seeing speech. Frequencies in the range of the voice pitch (approximately between 70 and 300 Hz) can be perceived by vibrotactile perception (Cholewiak & Collins, 1991). When hearing loss is profound, hearing aids must operate at high output levels that result in perceptible mechanical vibration (Bernstein, Tucker, & Auer, 1998). Boothroyd and Cawkwell (1970; see also Nober, 1967) studied the problem of distinguishing vibrotactile from auditory perception in adolescents with hearing losses. They found that sensation thresholds below 100 dB HL for frequencies as high as 1000 and even 2000 Hz might be attributable to detection of mechanical rather than acoustic vibration.

Perception of information for voicing might be obtained via a hearing aid through mechanical stimulation of the skin and might account for why some individuals with profound hearing losses obtain benefit from their hearing aids when communicating via speech. That voicing information can combine effectively with lipreading has been demonstrated in a number of studies. For example, Boothroyd, Hnath-Chisolm, Hanin, and Kishon-Rabin (1988) presented an acoustic signal derived from the voice pitch in combination with sentences presented visually to hearing participants. The mean visual-only sentence score was 26% words correct, and the audiovisual sentence score was 63%. Furthermore, we and others have demonstrated, using custom vibrotactile devices, that lipreading can be enhanced when voice fundamental frequency information is presented as vibration patterns on the skin, although the vibrotactile studies have generally failed to produce the same impressive gains obtained with analogous auditory signals and hearing participants (Auer, Bernstein, & Coulter, 1998; Eberhardt, Bernstein, Demorest, & Goldstein, 1990; Boothroyd, Kishon-Rabin, & Waldstein, 1995).

Summary and Conclusions

Speech information can withstand extreme degradation and still convey the talker's intended message. This fact explains why severe or profound hearing loss does not preclude perceiving a spoken language. Studies reviewed above suggest that listeners with hearing loss can profit from even minimal auditory information, if it is combined with visual speech information. Some individuals with profound hearing loss are able to perform remarkably well in auditory-only conditions and/or in visual-only conditions. However, the performance level that is achieved by any particular individual with hearing loss likely depends on numerous factors that are not yet well understood, including when their hearing loss occurred, the severity and type of the loss, their family linguistic environment, and their exposure to language (including their relative reliance on spoken vs. manual language).

Early studies of speech perception in hearing people focused on perception of the segmental consonants and vowels. More recently, research has revealed the importance of perceptual processes at the level of recognizing words. The studies reviewed above suggest the possibility that factors at the level of the lexicon might interact in complex ways with specific hearing loss levels. A complete understanding of the effectiveness of speech perception for individuals with hearing loss will require understanding relationships among the configuration of the hearing loss, the ability to amplify selected frequency regions, and the distinctiveness of words in the mental lexicon. These complex relationships will, in addition, need to be considered in relationship to developmental factors, genetic predispositions, linguistic environment, linguistic experience, educational and training opportunities, and cultural conditions.

Notes

1. The terms "lipreading" and "speechreading" are sometimes used interchangeably and sometimes used to distinguish between, respectively, visual-only speech perception and audiovisual speech perception in people with hearing losses. We have used both terms for visual-only speech perception. In this chapter, "lipreading" refers to perception of speech information via the visual modality.
2. The place distinction concerns the position in

the vocal tract at which there is critical closure during consonant production. For example, /b/ is a bilabial due to closure of the lips, and /d/ is a dental due to the closure of the tongue against the upper teeth. Manner concerns the degree to which the vocal tract is closed. For example, /b/ is a stop because the tract reaches complete closure. But /s/ is a fricative because air passes through a small passage. Voicing concerns whether or not and when the vocal folds vibrate. For example, /b/ is produced with vocal fold vibration almost from its onset, and /p/ is produced with a delay in the onset of vibration.

3. This correlation could have arisen because, at Gallaudet University, students who used their hearing aids more frequently were also more reliant on speech communication. That is, hearing aid use was a proxy in this correlation for communication preference/skill.

4. A phoneme is a consonant or vowel of a language that serves to distinguish minimal word pairs such as /b/ versus /p/ in "bat" versus "pat."

References

Auer, E. T., Jr. (2002). The influence of the lexicon on speechread word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*, *9*, 341–347.

Auer, Jr., E. T. & Bernstein, L. E. (1997). Speechreading and the structure of the lexicon: Computationally modelling the effects of reduced phonetic distinctiveness on lexical uniqueness. *Journal of the Acoustical Society of America*, *102*(6), 3704–3710.

Auer, E. T., Jr., & Bernstein, L. E. (2002). *Estimating when and how words are acquired: A natural experiment examining effects of perceptual experience on the growth the mental lexicon*. Manuscript submitted.

Auer, E. T., Jr., Bernstein, L. E. & Coulter, D. C. (1998). Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals. *Journal of the Acoustical Society of America*, *104*, 2477–2489.

Auer, Jr., E. T., Bernstein, L. E., & Tucker, P. E. (2000). Is subjective word familiarity a meter of ambient language? A natural experiment on effects of perceptual experience. *Memory & Cognition*, *28*(5), 789–797.

Berger, K. W. (1972). Visemes and homophenous words. *Teacher of the Deaf*, *70*, 396–399.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (1998). What makes a good speechreader? First you have to find one. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye (II): The psychology of speechreading and auditory-visual speech* (pp. 211–228). East Sussex, UK: Psychology Press.

Bernstein, L. E. Demorest, M. E., & Tucker, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*, 233–252.

Bernstein, L. E., Tucker, P. E., & Auer, E. T., Jr. (1998). Potential perceptual bases for successful use of a vibrotactile speech perception aid. *Scandinavian Journal of Psychology*, *39*(3), 181–186.

Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *27*, 134–143.

Boothroyd, A., & Cawkwell, S. (1970). Vibrotactile thresholds in pure tone audiometry. *Acta Otolaryngologica*, *69*, 381–387.

Boothroyd, A., Huath-Chisolm, T., Hanin, L., & Kishon-Rabin, L. (1988). Voice fundamental frequency as an auditory supplement to the speechreading of sentences. *Ear & Hearing*, *9*, 306–312.

Boothroyd, A., Kishon-Rabin L., & Waldstein, R. (1995). Studies of tactile speechreading enhancement in deaf adults. *Seminars in Hearing*, *16*, 328–342.

Breeuwer, A. & Plomp, R. (1984). Speech reading supplemented with frequency-selective sound-pressure information. *Journal of the Acoustical Society of America*, *76*, 686–691.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington, IN: Indiana University.

Ching, T. Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *Journal of the Acoustical Society of America*, *103*, 1128–1139.

Cholewiak, R., & Collins, A. (1991). Sensory and physiological bases of touch. In M. A. Heller & W. Schiff (Eds.), *The psychology of touch*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., & Fausti, S. A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear & Hearing*, *22*, 1–13.

Dunn, L. M. & Dunn, L. M (1981). *Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.

Eberhardt, S. P., Bernstein, L. E., Demorest, M. E., Goldstein, M. H. (1990). Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency. *Journal of the Acoustical Society of America*, *88*, 1274–1285.

Fisher, C. G. (1968). confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, *11*, 796–804.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281.

Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, *71*, 105–110.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, *103*, 2677–2690.

Howes, D. H. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, *29*, 296–305.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 245–294.

Lamoré, P. J. J., Huiskamp, T. M. I., van Son, N.J.D.M.M., Bosman, AJ., & Smoorenburg, G. F. (1998). Auditory, visual and audiovisual perception of segmental speech features by severely hearing-impaired children. *Audiology*, *37*, 396–419.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*, 148–167.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*, 187–196.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. (Research on Speech Perception, Technical Report No. 6). Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P. A., Goldinger, S.D., Auer, E.T., Jr., & Vitevitch, M.S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62(3)*, 615–625.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1–36.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122–147). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148–172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. (1992) Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: Bradford Books.

Mattys, S., Bernstein, L. E. & Auer, E. T., Jr., (2002). Stimulus-based lexical distinctiveness as a general word recognition mechanism. *Perception & Psychophysics*, *64*, 667–679.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18* 1–86.

McEvoy, C., Marschark, M., & Nelson, D. L. (1999). Comparing the mental lexicons of deaf and hearing individuals. *Journal of Educational Psychology*, *19*, 312–320.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*, 338–352.

Mogford, K. (1987). Lip-reading in the prelingually deaf. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 191–211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *Journal of the Acoustical Society of America*, *73*, 2134–2144.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116–133.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*, 3241–3254.

Nitchie, E. B. (1916). The use of homophenous words. *Volta Review*, *18*, 83–85.

Nober, E. H. (1967). Vibrotactile sensitivity of deaf children to high intensity sound. *Larynoscope*, *78*, 2128–2146.

Norris, D. (1994). Shortlist: A connectionist model of continuous word recognition. *Cognition*, *52*, 189–234.

Owens, E., & Blazek, B. (1985). Visemes observed by hearing impaired and normal hearing adult viewers. *Journal of Speech and Hearing Research*, *28*, 381–393.

Remez, R. E. (1994) A guide to research on the perception of speech. In: M. Ann Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 145–172). San Diego, CA: Academic Press.

Rönnberg, J. (1995). Perceptual compensation in the deaf and blind: Myth or reality? In R. A. Dixon & L. Bäckman (Eds.), *Compensating for psychological deficits and declines* (pp. 251–274). Mahwah, NJ: Lawrence Erlbaum Associates.

Rönnberg, J., Samuelsson, S., & Lyxell, B. (1998). Conceptual constraints in sentence-based lipreading in the hearing-impaired. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye: II. The psychology of speechreading and auditory-visual speech* (pp. 143–153). East Sussex, UK: Psychology Press.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215.

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I. G. Mattingly & M. Studert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 117–137). Hillsdale, NJ: Lawrence Erlbaum Associates.

Takayanagi, S., Dirks, D. D., & Moshfegh, A. (in press). Lexical and talker effects on word recognition among native and non-native normal and hearing-impaired listeners. *Journal of Speech, Language, Hearing Research*, *45*, 585–597.

Turner, C. W., & Brus, S. L. (2001). Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, *109*, 2999–3006.