RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Use of Partial Stimulus Information by Cochlear Implant Patients and Normal-Hearing Listeners in Identifying Spoken Words: Some Preliminary Analyses[1]

Lorin Lachs, Jonathan W. Weiss and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

[1] This research was supported by NIH T32 Training Grant DC00012 to Indiana University.

[2] Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Use of Partial Stimulus Information by Cochlear Implant Patients and Normal-Hearing Listeners in Identifying Spoken Words: Some Preliminary Analyses

Abstract. An error analysis of the word recognition responses of cochlear implant patients and normal-hearing listeners was conducted to determine the types of partial information used by these two populations when they identify spoken words under auditory-alone and audiovisual conditions. The results revealed that different types of partial information are used by the two groups in identifying spoken words under audio-alone or audiovisual presentation. Different types of partial information are also used in identifying words with different lexical properties. However, there were no significant interactions with hearing status, indicating that cochlear implant patients and normal-hearing listeners identify spoken words in a similar manner. The information available to patients with cochlear implants preserves much of the partial information necessary for accurate spoken word recognition.

Introduction

Cochlear implants are surgically inserted prosthetic devices that directly interface with the cochlea, providing electrical stimulation to the auditory nerve and thereby restoring hearing to those who had lost it. Although many users of the implants do well on standard outcome measures of word recognition, others receive less benefit from their devices (Pisoni, 1999; Pisoni, 2000). In order to improve cochlear implant (CI) design and use, it is necessary to determine the factors that give rise to this extensive variation. One possible source of variation lies in the sensitivity of cochlear implant users to the information necessary for accurate speech perception. In addition, there may be extensive individual differences in the process by which CI users utilize this sensory information during the process of spoken word recognition (Kirk, 2000; Pisoni, 2000; Pisoni, Cleary, Lachs, & Kirk, 2000). By comparing the performance of normal-hearing listeners and cochlear implant patients, we can examine the similarities and differences in the information each group perceives and the ways in which they use that information for spoken word recognition.

Several factors play well-established and important roles in speech perception and spoken word recognition. For example, although speech perception seems to be an inherently auditory process, a growing body of research has shown that *visual* information about speech can also be informative. In their pioneering study, Sumby and Pollack (1954) reported that the intelligibility of spoken words is enhanced when listeners are presented with both auditory and visual information, compared to auditory-only conditions. The addition of visual information can result in performance gains that are equivalent to an increase in signal-to-noise ratio of +15dB (Erber, 1969; Middleweerd & Plomp, 1987; Rosenblum & Saldaña, 1996; Sumby & Pollack, 1954). Under certain conditions, visual input may be a very important source of information about the speech signal, especially when acoustic information is degraded or unavailable. Several studies have shown that hearing impaired listeners (Erber, 1975; Massaro & Cohen, 1999; Tyler, Tye-Murray, & Lansing, 1988) and CI users (Lachs, Pisoni, & Kirk, in press; Tyler et al., 1997; Tyler, Opie, Fryauf-Bertschy, & Gantz, 1992) are also sensitive to the relationship between visual and auditory information, and make use of both sources during speech perception.

Another factor known to influence speech perception in normal-hearing listeners is talker variability. In clinical settings, cochlear implant patients frequently report better understanding of familiar voices, such as those of their spouses and other family members, than unfamiliar voices. Indeed, a listener's familiarity with the specific details of a talker's voice has been shown to improve speech intelligibility scores (Nygaard & Pisoni, 1998). Other studies have found processing costs associated with perceiving speech when it is produced by multiple talkers, as compared to speech produced by a single talker (Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989). These findings suggest that dealing with talker variability is a resource demanding process, and as such, plays an important role in speech perception and spoken word recognition.

Finally, numerous studies have demonstrated that the lexical properties of words affect speech perception under auditory-only presentation conditions (Luce & Pisoni, 1998). Frequency of occurrence in the language, neighborhood density, and average neighborhood frequency all affect recognition performance (Luce & Pisoni, 1998). Word frequency is a measure of how often a particular word is used in language. Neighborhood density refers to the number of words that sound similar to a target word. One way this is estimated is by counting the number of words that differ from a target word by one phoneme. This measure can be used as an index of phonological similarity in local regions of the lexicon. Neighborhood frequency is the average word frequency of all the words in a given phonological neighborhood. In addition to auditory-only speech perception, there is some recent evidence showing that these lexical factors also affect visual-alone speech perception (Auer & Bernstein, 1997; Lachs, 1999) and audiovisual speech perception (Kirk, Pisoni, & Osberger, 1995; Lachs et al., in press). In order to examine the effects of these lexical factors on word recognition, “easy” and “hard” words can be selected using the theoretical framework developed in the Neighborhood Activation Model (Luce, 1986; Luce & Pisoni, 1998). Easy words have a high frequency of occurrence, low neighborhood density, and low neighborhood frequency. The combination of these variables makes them relatively “easy” to perceive quickly and accurately. In contrast, hard words have the opposite characteristics: low frequency of occurrence, high neighborhood density, and high neighborhood frequency, resulting in these words being perceived more slowly and less accurately. The easy/hard lexical distinction - the most extreme conditions formed by the orthogonal combination of word frequency, neighborhood density, and neighborhood frequency - is a useful tool for examining speech perception performance under different presentation conditions.

In a recent paper, Kaiser, Kirk, Pisoni, and Lachs (2000) examined the spoken word recognition skills of cochlear implant patients and a group of normal-hearing listeners as a function of lexical discriminability, presentation mode, and number of speakers. All stimuli were isolated English monosyllabic words, presented under three conditions: audiovisual (AV), auditory-alone (A), and visual-alone (V). In addition, the words were grouped according to their lexical discriminability into two classes, Easy and Hard. In order to equate performance levels across the hearing groups, normal-hearing listeners heard auditory-alone and audiovisual stimuli in speech spectrum noise at a signal-to-noise ratio of –5 dB SPL. During the course of the experiment, listeners were presented with several lists of stimuli and asked to repeat the word being spoken. These lists varied according to the third factor included in the design, with some lists spoken by a single talker and other lists spoken by multiple talkers.

Several intriguing discoveries were made. First, Kaiser et al. confirmed that both normal-hearing listeners and cochlear implant patients correctly perceived spoken words with the greatest accuracy under audiovisual presentation. As expected, accuracy was lower under audio-alone presentation and even worse under visual-alone presentation conditions. In addition, Kaiser et al. found decreased word recognition accuracy for multiple-talker lists relative to single-talker lists. Confirming that, like normal-hearing listeners, CI patients also incur processing costs during the presentation of multiple-talker lists. Finally, Kaiser et al. found that both normal-hearing listeners and CI patients identified lexically easy words more accurately than lexically hard words under auditory-only, visual-only and audiovisual presentation conditions.

The results reported by Kaiser et al. demonstrate that CI patients are affected by many of the same factors that affect normal-hearing listeners during speech perception. In order to investigate potential differences more closely, Kaiser et al. also performed an analysis on the errors made by normal-hearing listeners and CI patients. Every error response was analyzed in order to determine if it was a phonological neighbor of the target word. The errors for both groups of listeners showed that under audiovisual conditions, a higher proportion of errors came from the target word's phonological neighborhood than under audio-alone conditions. This pattern suggests that responses to audiovisual stimuli, even when incorrect, were more accurate than responses to audio-alone stimuli. Because there were no overall differences in the error patterns between the two groups, the error analysis indicates that CI users are making use of partial information for word identification, resulting in responses that are phonetically similar to the target word, just as normal-hearing listeners do.

In order to examine the detailed characteristics of the partial information that cochlear implant patients used in the Kaiser et al. study, we performed several additional error analyses using broad phonetic categories. Broad coding eliminates distinctions between phonemes along particular perceptual dimensions and preserves distinctions along others (Huttenlocher & Zue, 1984; Miller & Nicely, 1955; Shipman & Zue, 1982). For example, the phonemes /p/, /b/, /m/, /β/, /ɸ/, /t/, /d/, /s/, /z/ and /n/ can be grouped together into two larger categories, /p b ɸ β m/ and /t d s z n/ by eliminating distinctions between phonemes based on their manner of articulation and by preserving distinctions based on place of articulation. In this case, one broad category consists of segments with bilabial place of articulation and the other consists of segments with alveolar place of articulation, but both contain segments that are stops, fricatives or nasals. The idea here is that broad coding allows the investigator to determine which featural distinctions are perceived and which ones are missed based on partial information. To continue with the example, if response accuracy in a set of trials increased after transcription with the broad categories outlined above, then one could conclude that responses were made based on perceived place of articulation, but not on perceived manner of articulation. This technique could be extremely useful in determining the type and quality of information transmitted by the cochlear implant under a variety of conditions.

Summerfield (1987) described the "Visual: Place, Auditory: Manner" (VPAM) model of audiovisual speech perception, which is a rough approximation of the types of information that can be obtained from the auditory and visual sensory modalities. VPAM is based on the assumption that during audiovisual speech perception, the place of articulation of an utterance is obtained through visual information and the manner of articulation is obtained through auditory information. Of course, this is an extremely simplified account of the actual process of audiovisual speech perception. However, in general, such a model is consistent with experimental evidence about the perceptual confusability of phonemes under audio- and visual-alone conditions (Summerfield, 1987; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). In general, phonemes that are highly confusable under auditory-alone conditions tend to be highly distinct under visual-alone conditions, and vice-versa. Confusions made under audio-alone conditions tend to be along the place of articulation dimension. For example, /f / and /θ/ are the most confusable phonemes in auditory noise (Miller & Nicely, 1955; Summerfield, 1987). Note that both are unvoiced fricatives, but differ in their place of articulation (/f/ is labiodental, /θ/ is dental). Visually, however, these two phonemes are highly distinct. In contrast, the phonemes /b/ and /m/ are virtually indistinct under visual-alone conditions, but are highly distinct under audio-alone conditions. To a rough approximation, this pattern is observed for the relationships among all phonemes: place distinctions are more easily made with visual information, while manner distinctions are more easily made with auditory information (Summerfield, 1987).

The present broad coding analysis was carried out to examine the patterns of errors made under various presentation conditions by cochlear implant patients and normal-hearing listeners to reveal both similarities and differences in spoken language processing for the two groups of listeners. We used two different broad categorization methods to represent the use of place and manner information during speech perception. In order to examine place of articulation, we broad coded the target and response words using eight categories based on the International Phonetic Association's places of articulation, collapsing across manner distinctions. The data were also scored using the broad categories described by Shipman and Zue (1982). In this classification method, six broad categories roughly approximate the different manners of articulation in English. Each target-response pair was examined using both of these broad coding methods to examine how partial phonetic information is used in an open-set word recognition task.

Method

Participants

Details of the methodology used for data collection in Kaiser et al. are provided here for convenience. Twenty postlingually deafened adult users of cochlear implants and nineteen normal-hearing adults participated in the original study. The hearing impaired adults had profound bilateral sensorineural hearing loss and a mean age of 50 years. All patients had more than six months of experience using their cochlear implant device and were recruited from the clinical population at Indiana University School of Medicine. The normal-hearing adults had a mean age of 40 years and were recruited from staff and students at Indiana University and the associated campuses.

Materials and Equipment

Six pairs of lexically-balanced word lists were formed from the Hoosier Audiovisual Multitalker Database (HAVMD), a digital database of audiovisual recordings of eight talkers speaking isolated monosyllabic English words (Lachs & Hernández, 1998; Sheffert, Lachs, & Hernández, 1996). Stimuli in the HAVMD were digitized at 640 x 480 resolution at 30 frames per second (fps). Audio tracks in the stimuli were digitized at 22 kHz with 16-bit resolution. In order to examine the effects of lexical discriminability, each list contained thirty-six words of which half were lexically hard and half were lexically easy. Lexical discriminability was determined by examination of the lexical characteristics of the 20,000 words in *Webster's Pocket Dictionary* (Nusbaum, Pisoni, & Davis, 1984). To examine the effects of talker variability, each pair of lists contained the same words but had a different number of talkers producing the words. One list in a pair was created using only a single talker; the other list was created using six different talkers each producing six words for the list. Using data obtained from normal-hearing adult listeners under visual-only presentation (Lachs & Hernández, 1998), the lists were then equated so that visual intelligibility was balanced across the various experimental conditions.

Each subject was tested in an IAC single-walled sound-treated booth. A PowerWave 604 (Macintosh compatible) computer with a Targa 2000 video board was used to present the digitized audiovisual stimuli. Each listener was presented with six different lists. Because the present analyses focused on responses to audio-alone and audiovisual stimuli, only four of these lists were used in the present analysis. A detailed analysis of the visual-only responses can be found in Kaiser et al. Each list contained eighteen lexically easy words and eighteen lexically hard words. A single talker produced two of the lists and groups of multiple talkers produced the other two. Within each talker condition, one list was presented in the audiovisual mode and the other list was presented in the audio-only mode. For the cochlear implant patients, each speech token was presented at 70 dB SPL (C weighted). Normal-hearing listeners were tested in speech spectrum noise at a –5 dB signal-to-noise ratio. All subjects were asked to repeat the word that was presented on each trial. The response to each speech token was recorded on-line by the experimenter, who typed the response into a file containing all the responses for a particular subject.

For the current analysis, all responses to all stimuli were compiled into one of four text files according to the Presentation format (audiovisual or audio alone) and Talker (single or multiple) conditions. The resulting four text files contained the target words and all the responses to those words. These text files were fed into a DECtalk DTC03 Text-to-Speech System, configured such that it could output an ASCII-based phonemic transcription of each target word and response (Bernstein, Demorest, & Eberhardt, 1994).

Place			"Zue"		
Name	Real	Random	Name	Real	Random
Vowels	u ju ʊ ɚ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w	u ju ʊ ɚ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ	Vowels and syllabic consonants	u ju ʊ ɚ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w ju l̩ m̩	u ju ʊ ɚ oʊ aʊ ɪ i ei ɛ æ ɔi ɔ ai ə ɑ ʌ w ju l̩ m̩
Bilabial	b p m m̩	p d ʒ ŋ n	Stops	p t k b d g	p θ w l ɹ b g v
Labiodental	f v	t ʃ ɹ m̩ z	Nasals	m n ŋ	t ʃ t͡s z m n
Dental	ð θ	k θ j l b	Strong fricatives	f θ h v ð	k j ð
Velar	k h ŋ g	f s ð m	Weak fricatives	s ʃ t͡s ʒ d͡ʒ z	f s h ʒ n
Alveolar	s l l̩ ɹ t d z n n	t͡s w d͡ʒ	Glides and semi-vowels	w j l ɹ	d d͡ʒ ŋ
Postalveolar and affricates	ʃ t͡s ʒ d͡ʒ	h l̩			
Palatal	j	g v n			

Table 1. Broad categories used in the present analysis. The "Real" columns outline categories patterned in a principled manner, as described in the text. The "Random" columns outline categories to which phonemes were randomly assigned.

The phonemic transcriptions of the four text files were then recoded using two "real" ("place" and "Zue") and two "random" ("random place", and "random Zue") broad coding methods. Table 1 shows the assignment of each phoneme to a broad category for each of the four coding methods. The method used to broad code by "place" was patterned after the International Phonetic Alphabet. All speech sounds in the target words and responses were classified according to their place of articulation. Of the speech sounds used in the target words and responses, only seven places of articulation were represented: 1) bilabial, 2) labiodental, 3) dental, 4) alveolar, 5) post-alveolar, 6) palatal, and 7) glottal. All vowels were broad coded into an eighth group.

The method used to broad code by "Zue" was constructed by using the 6-way classification of phonemes proposed in Shipman and Zue (1982). This classification scheme was chosen for two reasons. First, there is empirical evidence that such a broad coding scheme can preserve much of the information in the lexicon by maintaining a large proportion of unique words. Second, the classification system corresponds roughly to a classification system based on the manner of articulation of speech sound. The six resulting broad categories were defined as follows: 1) vowels and syllabic consonants, 2) stops, 3) nasals, 4) strong fricatives, 5) weak fricatives, and 6) glides and semi-vowels.

	Phonetic	Place		Phonetic	Zue	
		real	random		real	random
Target	bæt	/bilabial/ V /alveolar/	/4/ V /3/	kæt	/stop/ V /stop/	/4/ V /3/
Response	mæd	/bilabial/ V /alveolar/	/5/ V /2/	pæk	/stop/ V /stop/	/2/ V /4/
Correct?	NO	YES	NO	NO	YES	NO

Table 2. Scoring methods used for target-response pairs. The "Target" row contains the various transcriptions of a target word. The "Reponse" row contains the various transcriptions of a response to the target word above it. The "Correct?" row shows whether the target-response pair was graded correct or incorrect under the relevant coding method. The columns denote the different coding methods. Numbers in the "random" columns denote the set to which the phoneme was randomly assigned, as outlined in Table 1.

We expected that by loosening the criterion for a correct response, overall accuracy would increase. However, there are two possible sources for this improvement in accuracy. First, accuracy might increase because the broad categories used might more accurately reflect the information perceived by the listener. For example, it is well known that the distinction between /b/, /p/, and /m/ is very hard to make when speechreading (Summerfield, 1987). It is common in analyses of speechreading data, therefore, to group these phonemes into a larger equivalence class, or "viseme," and count as correct any responses that substitute one member of the class for another (Bernstein et al., 1994; Walden et al., 1977). Accuracy scores improve because it is relatively easy to distinguish bilabially articulated phonemes from other phonemes, while it is more difficult to make distinctions based on voicing and nasality during speechreading. The drawback to this approach, however, is that the odds of randomly picking a correct segment increase. If there are only six response categories, the chance of randomly choosing the correct feature is much greater than if there are more than 40 response categories (as there are with phonemes).

In order to control for improvements in accuracy due to chance, two "random" broad coding methods were constructed. The "random place" and "random Zue" broad coding methods contained the same number of broad categories as their "real" counterparts. However, for each random coding method, each phoneme in the target set was randomly assigned to a broad category. Using the random coding as a benchmark, we can then determine how much improvement is due to the use of partial information in the stimulus and how much is due to just a decrease in the number of response categories.

The phonemic transcriptions of each of the target-response pairs, along with the broad-coded versions of the transcription, were analyzed using a custom-designed scoring program. Each target-response pair under each broad coding method was scored as correct if the target and response were identical. Table 2 shows two representative target-response pairs under each transcription and whether the target-response pair would be considered correct.

As discussed earlier, target words were presented under eight experimental conditions, based on three factors (Lexical Disciminability, Number of Talkers, and Presentation Modality) with two levels each. An index of broad coding enhancement (Y) was determined for each subject in each of the eight conditions. Y is therefore a measure of any possible improvement due to broad coding, normalized by the amount of gain that could have occurred. Y can also be conceptualized as the proportion of error responses that are scored as correct due to the broad transcription process. Y was calculated by the following formula:

$$Y = {(b-p)}/{(100-p)}$$

where “b” is the percent correct performance for a given broad coding method and “p” is the percent correct performance under phonetic transcription.

Results and Discussion

The enhancement score, Y, for each subject under each experimental condition was calculated and submitted as the dependent variable to a six-way (Number of Talkers, Presentation Mode, Lexical Discriminability, Hearing Group, Broad Categorization and Coding) ANOVA. The first four factors were from the original design. Talker had two levels: single talker (ST) and multiple talker (MT). The two levels of Presentation Mode were auditory-only (A) and audiovisual (AV). Lexical Discrimination also had two levels: Easy (E) and Hard (H). Finally, Hearing Group had two levels: normal hearing (NH) and cochlear implant (CI) patients. The last two factors were relevant to the present error analysis. The first was Broad Categorization, which had two levels: “place” and “Zue”. The second was Coding and had two levels: “real” and “random”.

There were no interactions between the Number of Talkers factor and the other factors in the analysis. The findings below are collapsed across the two levels of this factor.

Effects of Presentation Mode

Figure 1 shows the enhancement scores (Y) for responses to words presented under audiovisual or audio-alone conditions for “real” and “random” coding methods, separated by broad categorization and collapsed across Talker, Lexical Discriminability, and Hearing Group. The left panel shows the data from “place” transcription and the right panel shows the data from “Zue” transcription. Within each panel, the left set of bars shows audiovisual presentation data and the right set of bars shows audio-alone presentation data. Within each set of bars, the dark shaded bars show data from “real” coding and the open bars show data from “random” coding. The ANOVA revealed a significant three way interaction of Coding by Broad Categorization by Presentation, $F(1,38) = 26.503$, $MSE = 0.312$, $p < 0.001$, $\eta^2 = 0.411$. This interaction indicates that the relative effectiveness of each coding scheme, as compared with it’s random counterpart, was different based on the Presentation Modality. Furthermore, these three factors did not interact with Hearing Group; thus, the patterns below reflect the behavior of both normal-hearing listeners and cochlear implant patients. The three-way interaction was probed in depth by splitting the data along the Broad Categorization (Place vs. Zue) factor.

For target-response pairs transcribed by preserving place of articulation (i.e., for the data in the left-hand panel of Figure 1), we found a main effect of Coding, $F(1, 39) = 307.56$, $MSE = 0.546$, $p < 0.001$, $\eta^2 = 0.887$. Simple effects analysis revealed that relative difference scores in the “real” coding condition were greater than those in the “random” coding condition for both audiovisual, $F(1,39) = 208.133$, $p < 0.001$, $\eta^2 = 0.842$, and audio-alone, $F(1,39) = 25.335$, $p < 0.001$, $\eta^2 = 0.394$, presentations. In other words, regardless of the presentation modality, responses were correct more often under principled coding methods than they were under randomly constructed ones. This is not surprising given the fact that the principled coding methods are based on the selective collapsing and/or preservation of perceptually discriminable dimensions, whereas the random coding method follows no such constraints. It is based solely on the qualification that there be a set number of response categories.

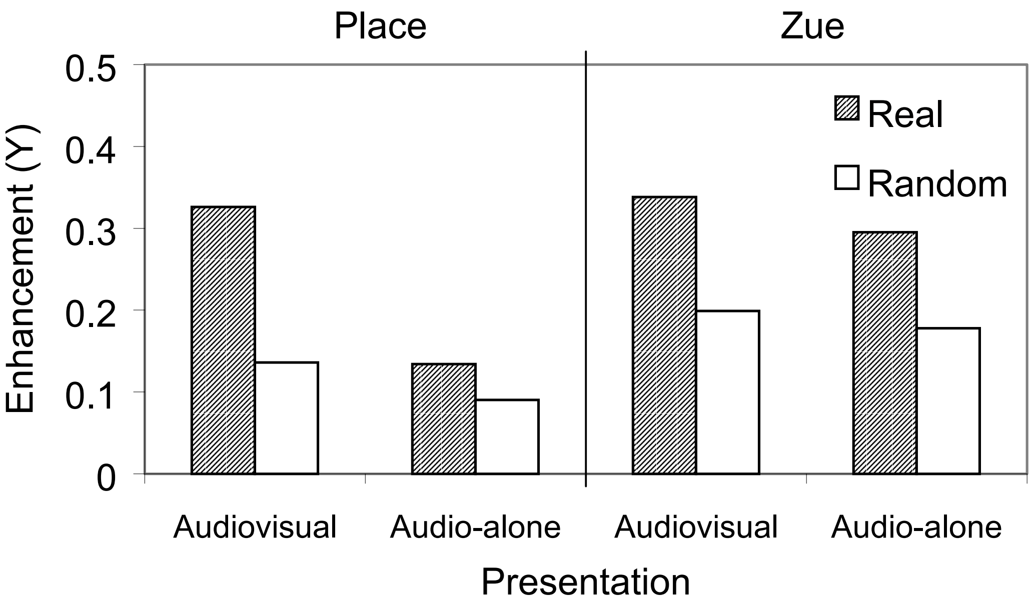

Figure 1. The relative difference scores for gains in accuracy due to broad transcription (Y) for responses to words presented under audiovisual or audio-alone conditions for "real" and "random" coding schemes separated by Broad Categorization. The data is collapsed across Talker, Lexical Discriminability, and Hearing Group.

More importantly, we also found a significant interaction between Presentation and Coding, $F(1,39) = 67.125$, $MSE = 0.215$, $p < 0.001$, $\eta^2 = 0.633$. It is clear from Figure 1 that, relative to the "random"-coding baselines, audiovisual responses benefited more from place-transcription than did audio-alone responses. Because Y represents the proportion of incorrect responses that became correct under transcription, this result indicates that in audiovisual conditions responses were closer to the target *even when they were phonetically incorrect* than they were under audio-alone conditions. In other words, audiovisual presentation elicited more accurate responses than did audio-alone presentation. Because this transcription method preserved distinctions by place of articulation, we can conclude that these "more accurate" responses to audiovisual stimuli were more accurate because they preserved the place of articulation of the segments in the target word. Even when subjects responded inaccurately in the audiovisual condition, their error responses were based closely on place of articulation: a perceptually salient dimension of visual speech. In other words, these subjects perceived partial information about the stimulus.

A slightly different picture emerges for target-response pairs transcribed with the "Zue" coding method (the right-hand panel of Figure 1). As with the "place" categorization, we observed a main effect of Coding, $F(1,39) = 116.604$, $MSE = 0.655$, $p < 0.001$, $\eta^2 = 0.749$. Simple effect analysis revealed that relative difference scores in the "real" coding condition were greater than those in the "random" coding condition for both audiovisual, $F(1,39) = 63.638$, $p < 0.001$, $\eta^2 = 0.62$, and audio-alone, $F(1,39) = 77.64$, $p < 0.001$, $\eta^2 = 0.666$, presentation. Again, this is not a surprising result, but is necessary to establish the validity of gains in accuracy due to the broad coding procedure.

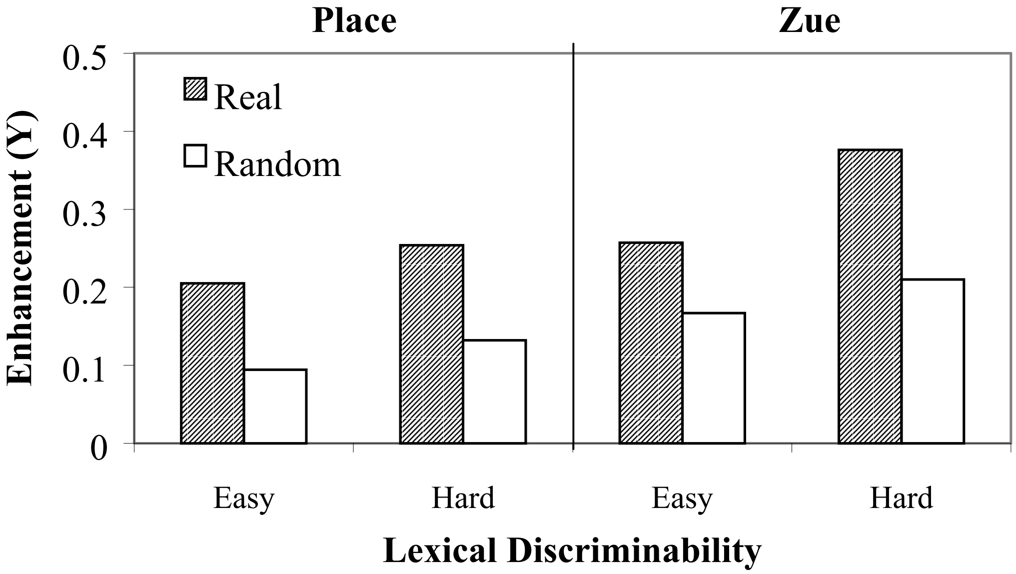

Figure 2. The relative difference scores of percent correct (Y) for responses to easy or hard words for "real" and "random" coding schemes separated by Broad Categorization. The data is collapsed across Talker, Presentation, and Hearing Group.

In contrast to the pattern observed with place transcription, the interaction of Coding by Presentation for responses transcribed with the "Zue" method was not significant, F(1,39) = 1.171, *MSE* = 0.0047, n.s. In general, there was no difference between the improvement in audiovisual scores and the improvement in audio-alone scores when they were classified according to the "Zue" method. This is interesting because it implies that, to a rough approximation, the responses elicited under either audiovisual or audio-alone presentation equally preserved the manner of articulation of the segments in the target word. Because the auditory stimulus was identical in the two Presentation conditions, this fact is not surprising; to a rough approximation, the manner of articulation is most perceptually salient via the acoustic modality.

Thus, we can conclude that the Place coding method accurately represented the kinds of partial information available under audiovisual presentation, as opposed to auditory-alone presentation, for both normal-hearing listeners and cochlear implant patients. The fact that these factors did not interact with Hearing Group indicates that both groups were equally sensitive to the additional place information available under audiovisual presentation.

Effects of Lexical Discriminability (Easy vs. Hard)

Figure 2 shows the enhancement scores (Y) for responses to easy or hard words for "real" and "random" coding methods separated by Broad Categorization. The data in Figure 2 is collapsed across Talker, Presentation, and Hearing Group. The left panel shows the data from "place" transcription and the right panel shows the data from "Zue" transcription. Within each panel, the left set of bars shows data from lexically easy target words and the right set of bars shows data from lexically hard target words. Within each set, the dark shaded bar shows data from "real" coding and the open bar shows data from "random" coding. The ANOVA showed a significant three way interaction of Coding by Broad

Categorization by Lexical Discriminability, $F(1,38) = 8.945$, $MSE = 0.0319$, $p < 0.01$, $\eta^2 = 0.191$. The interaction was probed by splitting the data along Broad Categorization.

For target-response pairs transcribed by place, we found a main effect of Coding, $F(1,39) = 307.562$, $MSE = 0.546$, $p < 0.001$, $\eta^2 = 0.887$. As above, the main effect of coding demonstrates that "real" scores were greater than the "random" scores, confirming that scores did not improve simply due to a reduction in the number of response categories.

We also found an effect of Lexical Discriminability, $F(1,39) = 8.472$, $MSE = 0.07426$, $p < 0.01$, $\eta^2 = 0.178$, indicating that incorrect responses (phonetically) to Hard words were more accurate than were incorrect responses to Easy words. Although this seems paradoxical at first glance, it is entirely consistent with the definition of Hard words. One of the components that defines a Hard word as Hard is that it has more phonetically similar neighbors than does an Easy word. Accuracy in this study is based on phonetic similarity. Numerically, a response has a better chance of being a neighbor of a Hard word than it does of being a neighbor of a hard word. Indeed, Kaiser et al. (2000) conducted a neighbor analysis of the present dataset and found that incorrect responses to Hard target words were more often within the neighborhood of the target word than were responses to Easy target words.

Interestingly, however, there was no interaction between Coding and Lexical Discriminability, $F(1,39) < 1$. Thus, relative to the "random" score baselines, the benefit gained by "place" transcription was not biased toward either easy or hard words. In other words, responses for Easy and Hard words did not differ in the extent to which they preserved the place of articulation of the segments in the target word.

However, for the "Zue" transcription, a significant interaction of Coding by Presentation was observed, $F(1,39) = 20.013$, $MSE = 0.058$, $p < 0.001$, $\eta^2 = 0.339$. These results indicate that relative to "random" score baselines, "Zue" broad categorization benefits hard words more than easy words. Therefore, subjects preserved the manner of articulation of the target words more often for Hard words than they did for Easy words. The interaction indicates that for Hard target words, responses were more accurate than responses to Easy words, even when they were phonetically incorrect. Because "Zue" transcription preserves distinctions by manner of articulation, we can conclude that the "more accurate" responses to Hard words were more accurate by virtue of the fact that they preserved the manner of articulation of the segments in the target word.

Again, there was no interaction between these three factors and Hearing Group, indicating that the use of partial information was consistent between groups. It is not surprising that the two hearing groups demonstrated similar effects of Lexical Discriminability: the cochlear implant patients tested were all post-lingually deafened adult speakers of American English. Presumably, for these patients, lexical structure developed normally before their hearing loss. To the extent that their learned knowledge about the similarity patterns among spoken words did not change much during the period between hearing loss and implantation, we can expect that effects of lexical structure on spoken word recognition might not diminish due to cochlear implantation.

Effects of Hearing Group (CI vs. NH)

Although there were no interactions between the factors we tested and Hearing Group, normal-hearing listeners and cochlear implant patients did differ in the overall extent to which they were sensitive to partial information. On average, the responses given by cochlear implant patients benefited more from broad categorization than those of normal-hearing listeners, $F(1, 38) = 4.1$, $MSE = 0.07$, $p = .05$, $\eta^2 = 0.098$. The mean enhancement score for cochlear implant patients was 0.25 ($SE = .02$), indicating that broad coding raised CI scores about 25% of the amount they could have improved. In contrast, the mean enhancement score for normal-hearing listeners was 0.20, indicating that normal hearing scores improved by 20% of their potential gain due to broad categorization. Thus, when cochlear implant patients made an error in identification, they were closer to the target than when normal-hearing listeners made an error.

This result could have been observed for two reasons. First, it should be noted that adding noise to the auditory signal for normal-hearing listeners is hardly a degradation of the signal that is analogous to that experienced by cochlear implant patients. The main effect of Hearing Group observed here might simply be due to a failure on the part of Kaiser et al.'s manipulation designed to equate relative performance levels across the conditions. Alternatively, the main effect may also be due to a learned ability on the part of cochlear implant patients to make better use of partial information than their normal-hearing counterparts do, due to their experience with degraded inputs.

The fact that the Hearing Group factor did not interact with any of the other factors known to effect normal-hearing listeners' spoken word recognition, however, demonstrates that post-lingually deafened adult cochlear implant patients are sensitive to the partial information available under varying conditions, and use this information in much the same way that normal-hearing listeners do.

Discussion

In this study we examined how normal-hearing listeners and cochlear implant patients differ in their use of partial information in the perception of words. All of the listeners were presented with spoken words under a number of different conditions that manipulated presentation modality, lexical status, and the number of talkers. The target-response pairs were broad coded allowing a more detailed investigation into the perception of spoken words. By examining the structure and patterns of incorrect responses, we were able to draw several conclusions about the partial information used by perceivers with normal hearing and cochlear implants.

Effects of Presentation

We found that responses preserved place of articulation when targets were presented audiovisually more often than when targets were presented audio-alone. This result was probably obtained because information regarding place of articulation is more perceptually robust and better specified visually in audiovisual stimuli than it is in audio-alone presentation (Summerfield, 1987), especially for normal-hearing listeners listening in noise and for cochlear implant patients. Of course, as shown in Figure 1, audio-alone scores also benefited from "place" transcription. Obviously, some information regarding place is contained within the audio-alone signal perceived by the subject. However, information regarding place may either be incomplete or partially degraded during audio-alone presentation relative to the information available concerning place during audiovisual presentation, especially for cochlear implant users and normal-hearing listeners in noise.

In contrast, the availability of perceptual information regarding manner of articulation did not appear to change between audiovisual and audio-alone presentation. Responses were similar in the extent to which they were sensitive to manner under both presentation conditions. This suggests that perceptual information about manner in audio-alone presentation does not differ much from the information available in audiovisual presentation. Apparently, the addition of a visual signal does not affect (to a great extent) the information relevant to the manner of articulation contained within the auditory signal alone for either normal-hearing listeners or cochlear implant patients.

These results also indicate that cochlear implant and normal-hearing subjects do not perceive audiovisual or audio-alone perceptual information differently, demonstrating that basic processes of spoken word recognition may be common among normal-hearing subjects and users of cochlear implants.

Effects of Lexical Discriminability

Responses to both easy and hard target words preserved the place of articulation of the target to the same degree. In contrast, there *were* differences in the extent to which responses to easy and hard words preserved manner of articulation. Specifically, responses to hard target words preserved the manner of articulation of targets more than responses to easy target words. Such a result is consistent with an explanation based on the distribution of lexical distinctions across the lexicon. It seems reasonable to assume that, for any given target word, the proportion of neighbors that differ by a particular phonetic distinction would be constant. This is of course an empirically testable assumption. If it were true, that would imply that the likelihood of a target word being identified as a neighbor by a particular distinction would be calculable. The results presented above would then be consistent with a scenario in which the proportion of neighbors that differ by manner, for each word in the entire lexicon, is higher than the proportion of neighbors that differ by place, but only in dense neighborhoods! In other words, the phonological neighbors of hard target words are more similar in regards to manner than the phonological neighbors of easy target words. An in-depth study of the distribution of neighborhood distinctions could test this hypothesis.

In this study we found that cochlear implant and normal-hearing subjects use partial information similarly in identifying spoken words. This implies that the loss of hearing does not significantly change the partial information used during speech perception or the process by which spoken words are recognized. Thus, cochlear implant devices appear to allow the perception of partial information that is useful to speech perception under a variety of conditions.

References

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

Bernstein, L. E., Demorest, M. E., & Eberhardt, S. P. (1994). A computational approach to analyzing sentential speech perception: Phoneme-to-phoneme stimulus-response alignment. *Journal of the Acoustical Society of America, 95*, 3617-3622.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423 - 424.

Erber, N.P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*, 481-492.

Huttenlocher, D. P., & Zue, V. W. (1984). *A model of lexical access from partial phonetic information.* Paper presented at the The IEEE International Conference on Acoustics, Speech and Signal Processing, San Diego, CA.

Kaiser, A. R., Kirk, K. I., Pisoni, D. B., & Lachs, L. (2000). *Audiovisual speech perception of adult cochlear implant users: Integration, talker and lexical effects.* Manuscript in preparation.

Kirk, K.I. (2000). Challenges in the clinical investigation of cochlear implant outcomes. In J. K. Niparko, K. I. Kirk, N. K. Mellon, A. M. Robbins, D. L. Tucci, & B. S. Wilson (Eds.), *Cochlear Implants: Principles and Practices* (pp. 225 - 258). Philadelphia, PA: Lippincott, Williams and Wilkins.

Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing, 16*, 470 - 481.

Lachs, L. (1999). Use of partial stimulus information in spoken word recognition without auditory stimulation, *Research on Spoken Language Processing No. 23*. Bloomington, IN: Speech Research Laboratory, Indiana University Bloomington.

Lachs, L., & Hernández, L. R. (1998). Update: The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing Progress Report 22* (pp. 377 -388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., Pisoni, D. B., & Kirk, K. I. (in press). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing*.

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

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

Massaro, D. W., & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language, and Hearing Research, 42*, 21 - 41.

Middleweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold in noise. *Journal of the Acoustical Society of America, 82*, 2145 - 2147.

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

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words, *Research on Speech Perception Progress Report No. 10.* Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 355 - 376.

Pisoni, D. B. (1999). *Individual differences in the effectiveness of cochlear implants.* Paper presented at the 138th Meeting of the Acoustical Society of America, Columbus, OH.

Pisoni, D. B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear & Hearing, 21*, 70 - 78.

Pisoni, D. B., Cleary, M., Lachs, L., & Kirk, K. I. (2000, November, 2000). *Individual differences in effectiveness of cochlear implants in prelingually deaf children.* Paper presented at the American Speech-Language Hearing Association Annual Convention, Washington, D. C.

Rosenblum, L.D., & Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 318-331.

Sheffert, S. M., Lachs, L., & Hernández, L. R. (1996). The Hoosier Audiovisual Multitalker Database, *Research on Spoken Language Processing No. 21* (pp. 578 - 583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Shipman, D. W., & Zue, V. W. (1982). *Properties of large lexicons: Implications for advanced isolated word recognition systems.* Paper presented at the IEEE 1982 International Conference on Acoustics, Speech and Signal Processing.

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3 - 51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tyler, R. S., Fryauf-Bertschy, H., Gantz, B. J., Kelsay, D. M. R., Tyler, R. S., Woodworth, G. G., & Parkinson, A. (1997). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology-Head and Neck Surgery, 117*, 180 -187.

Tyler, R. S., Opie, J. M., Fryauf-Bertschy, H., & Gantz, B. J. (1992). Future directions for cochlear implants. *Journal of Speech-Language Pathology and Audiology, 16*, 151 - 163.

Tyler, R. S., Tye-Murray, N., & Lansing, C. R. (1988). Electrical stimulation as an aid to speechreading. Special Issue: New reflections on speechreading. *Volta Review, 90*, 119-148.

Walden, B.E., Prosek, R.H., Montgomery, A.A., Scherr, C.K., & Jones, C.J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20*, 130 - 145.

This page left blank intentionally.