

Stimulus-based lexical distinctiveness as a general word-recognition mechanism

SVEN L. MATTYS, LYNNE E. BERNSTEIN, and EDWARD T. AUER, JR.
House Ear Institute, Los Angeles, California

Word recognition is generally assumed to be achieved via competition in the mental lexicon between phonetically similar word forms. However, this process has so far been examined only in the context of auditory phonetic similarity. In the present study, we investigated whether the influence of word-form similarity on word recognition holds in the visual modality and with the patterns of visual phonetic similarity. Deaf and hearing participants identified isolated spoken words presented visually on a video monitor. On the basis of computational modeling of the lexicon from visual confusion matrices of visual speech syllables, words were chosen to vary in visual phonetic distinctiveness, ranging from visually unambiguous (lexical equivalence class [LEC] size of 1) to highly confusable (LEC size greater than 10). Identification accuracy was found to be highly related to the word LEC size and frequency of occurrence in English. Deaf and hearing participants did not differ in their sensitivity to word LEC size and frequency. The results indicate that visual spoken word recognition shows strong similarities with its auditory counterpart in that the same dependencies on lexical similarity and word frequency are found to influence visual speech recognition accuracy. In particular, the results suggest that stimulus-based lexical distinctiveness is a valid construct to describe the underlying machinery of both visual and auditory spoken word recognition.

The perceptual and cognitive mechanisms supporting spoken word recognition have been the object of ongoing research for several decades. An emerging consensus in modeling spoken word recognition is that lexical candidates compete for recognition as a function of their mutual form-based similarity (e.g., Forster, 1979; Marslen-Wilson, 1987; McQueen, Cutler, Briscoe, & Norris, 1995; Morton, 1979; Norris, 1994). Competition has been formalized in TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and the neighborhood activation model, or NAM (P. A. Luce, 1986; P. A. Luce & Pisoni, 1998; P. A. Luce, Pisoni, & Goldinger, 1990). For such models, words are recognized relationally, in the context of the other words in the mental lexicon. A word with few similar-sounding lexical neighbors—based on phoneme replacement, deletion, and addition, for example—is identified more easily than one in a dense region of the lexicon. The influence of phonetic similarity on lexical activation is supported by a large body of empirical evidence (e.g., Cluff & Luce, 1990; P. A. Luce, 1986; P. A. Luce & Pisoni, 1998; Marslen-Wilson & Warren, 1994) and is shown to hold when words are presented in noise (P. A. Luce & Pisoni, 1998).

Yet, despite the implicit assumption that this mechanism extends to the general arena of spoken word recognition, the evidence currently available is largely restricted to *auditory* spoken word recognition. The goal of the present study was to investigate the underlying machinery of the spoken word recognition process beyond the auditory modality and, thus, beyond the scope of the existing literature.

Research on auditory spoken word recognition limits understanding not only to a specific perceptual modality, but also to the patterns of phonetic similarity defined by that channel. For example, acoustic speech presented at reduced intensity or in a noisy background maintains voiced/voiceless distinctions but reduces the acoustic cues for place of articulation (Breeuwer & Plomp, 1985; Grant & Braida, 1991; Grant & Walden, 1996; Wang & Bilger, 1973). Thus, auditorily, *pack* (voiceless bilabial) is more similar to *tack* (voiceless alveolar) than it is to *back* (voiced bilabial). In contrast, visible speech (i.e., the talking face) tends to reduce the voicing contrast but transmit reliable optical cues for place of articulation. Here, *pack* is visually more similar to *back* than it is to *tack*. This example illustrates the fact that a word's competitor environment could change dramatically as a function of its stimulus condition.

In this study, we tested the hypothesis that word recognition is essentially a process of lexical discrimination of the target word from the words stored in the form-based mental lexicon, even when the speech input is visual—that is, when the patterns of phonetic similarity are generated by a different set of conditions. Evidence that visual spoken word recognition is influenced by stimulus-based lexical similarity would support the generalizability of the phenomenon across modalities and stimulus conditions.

This research was supported by Grant DC02107 from NIH/NIDCD under the direction of the PI, L.E.B. We thank Paula E Tucker and Sheri Hithe-Hultgren for their assistance in testing participants and Brian Chaney for his help with computer programming. Correspondence concerning this article should be addressed to S. L. Mattys, who is now in the Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, England (e-mail: sven.mattys@bris.ac.uk), or to L. E. Bernstein, Department of Communication Neuroscience, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057 (e-mail: lbernstein@hei.org).

The visual speech stimulus is phonetically impoverished relative to the auditory speech stimulus presented under favorable listening conditions. As a result, spoken word recognition is generally less accurate in the visual than in the auditory modality. A simple intuitive account of this difference could be that lexical access fails because phonemes are underspecified in the optical signal (e.g., /b/, /p/, and /m/ merge into a single phoneme equivalence class, or PEC—i.e., a group of phonemes with high perceptual confusability; Auer & Bernstein, 1997). Accordingly, lipreading performance would simply be a function of the average confusability of a word's individual phonemes. However, computational research provides evidence that stimulus-based lexical dissimilarity can reduce the problem of phonetic impoverishment in visual spoken word recognition (Auer & Bernstein, 1997; Iverson, Bernstein, & Auer, 1998). For example, even though /b/, /p/, and /m/ are visually similar, the word *bought* is a distinct visual spoken word form in the lexicon, because *pought* and *mought* are not valid lexical competitors in English. By contrast, the word *bat* is susceptible to lexical confusion because competitors such as *pat* and *mat* have their own lexical entry. Therefore, we hypothesized that the effects of PECs have to be considered in the context of the lexicon to be predictive of speech recognition accuracy.

The conjunction of PEC measurements and lexical data lead to the concept of lexical equivalence class (LEC) size (Auer & Bernstein, 1997), an index of lexical form-based similarity. For instance, the size of the LEC containing the word *bat* is greater than that containing *bought*—there are more word forms similar to *bat* than similar to *bought*. Thus, under the assumption that visual spoken word recognition is influenced by form-based distinctiveness, we expect LEC size to be a strong predictor of visual spoken word recognition accuracy above and beyond phonemic confusion. We refer to this as the *lexical distinctiveness hypothesis*.

It should be noted, however, that the concept of LEC is a convenient simplification of the problem of spoken words' visual similarity. Estimates of phonemic similarity, and, hence, estimates of lexical similarity can vary widely, depending on talker, lipreader, and situational characteristics (e.g., Jackson, 1988; Kricos & Lesner, 1982, 1985; Montgomery & Jackson, 1983) and, of course, on the confusability criterion chosen to cluster phonemes (Auer & Bernstein, 1997). LEC size, therefore, should be considered a tool for estimating the similarity of words, rather than an index of unalterable visual equivalence.

In the following experiment, we investigated the effect of lexical distinctiveness on word recognition by considering words from three classes of visual equivalence: words predicted to have no visual within-class competitors (LEC size of 1), words predicted to have a few competitors (LEC size of 2–6), and words predicted to have many competitors (LEC size of 10–60). In the context of LECs, *competitors* are defined as words that, computationally, are predicted to be perceptually identical or highly similar to

the target (e.g., *pat* and *mat* are competitors of *bat*). Thus, the LEC index acts as a threshold on predicted perceptual distinctiveness for words, particularly for lexical candidates that are theoretically *homophonous* with the target (i.e., visually highly similar but auditorily distinct, Berger, 1972; Nitchie, 1916). If visual spoken word recognition, like auditory word recognition, is accomplished relative to the form-based similarity of the other words of the lexicon, recognition accuracy should be inversely proportional to the LEC size of the test words.

Lexical influence on visual speech recognition is also likely to manifest itself through word frequency differences. Studies of auditory word recognition have shown repeatedly that high-frequency words are identified more easily than low-frequency words (Balota & Chumbley, 1984; Forster, 1976; Howes, 1954, 1957; Savin, 1963; Soloman & Postman, 1952). However, word frequency has never, to our knowledge, been examined explicitly in research on visual speech recognition. Whether the well-known frequency effect occurs for visual spoken word processing was a question for the present study. The frequency factor is of particular interest in the case of visually unique words (i.e., LEC size equal to one), because the phonetic distinctiveness of these words makes them most similar to their auditory counterparts. That is, in both cases, the stimulus theoretically provides sufficient information for unambiguous identification. Thus, according to our lexical distinctiveness hypothesis, lexical competition should be at its lowest level in the case of visually unique words. Frequency effects among these words would indicate that, regardless of the degree of distinctiveness, lexical activation through visual speech is carried out in a frequency-sensitive way, as it is the case in the auditory modality.

Finally, in the present study, words were also chosen among monosyllables and disyllables. The word length factor was motivated by the fact that a majority of the studies examining lexical similarity have been restricted to monosyllables. Monosyllables differ from longer words in several ways that could affect the ease with which they are identified. First, monosyllables are more frequent than words of other lengths. Moreover, because they have fewer segments, monosyllables have more potential to be phonetically similar to one another than longer words do. On the basis of confusion matrices obtained from visual nonsense syllable identification, Iverson et al. (1998) calculated that only 15% of all monosyllables are visually unique, whereas over 75% of all longer words are. On the other hand, multisyllables frequently contain reduced vowels, which could lower the visual intelligibility of the segments in the unstressed syllables. We controlled for word length to help us determine whether the effects of LEC size and word frequency come about equally despite these differences.

The focus of this study is unique not only because it aims to provide evidence for general modality-independent mechanisms involved in spoken word recognition, but also because, thus far, research on lipreading has been princi-

pally concerned with phonetic perception (via phoneme identification) and sentence comprehension (via transcription of words in sentences). The literature is notably scarce on how access to the word-form lexicon is achieved via visual perception of speech. Thus, the aim of the present experiment was also to provide insight into how optical phonetic information is processed in the absence of postlexical, sentential semantic context information.

All test words were presented to adults with normal hearing (NH) and adults with profound, prelingual hearing impairment (HI). The participants were prescreened for close-to-average or better lipreading ability. By testing both groups, our aim was to examine whether long-term experience relying on visual speech affects the mechanisms involved in visual spoken word recognition. Similar performance across groups would support similar visual spoken word recognition mechanisms despite different perceptual experience. By screening the participants for close-to-average or better lipreading ability within their reference population, our aim was to observe how lexical properties affect spoken word recognition in individuals whose visual spoken word processing system is comparable in terms of its overall effectiveness.¹

METHOD

Participants

NH and NI participants were screened for the following characteristics: (1) between 18 and 40 years of age, (2) no self-reported learning disabilities, (3) vision 20/30 or better in each eye, as determined with a standard Snellen chart, (4) self-reported use of English as the native language (including a manually coded form), and (5) better than half a standard deviation below the mean on a lipreading screening test, as referenced to the appropriate distribution of performance by deaf or hearing college-educated adults (Bernstein, Demorest, & Tucker, 1998). Specifically, the percentage of words correctly identified in 30 video-recorded sentences (Bernstein & Eberhardt, 1986) from the CID (Central Institute for the Deaf) Everyday Sentences (Davis & Silverman, 1970) was 29% among the NH participants and 47% among the HI participants. Normative data (Bernstein, Iverson, & Auer, 1997) indicate a mean of 22% (*SD* = 15.56) for NH participants and 44% (*SD* = 21.39) for HI participants. The participants were also selected to (6) have a vocabulary knowledge around the average level of their age group. All the participants were administered Form L of the Peabody Picture Vocabulary Test-Revised (PPVT; Dunn & Dunn, 1981). PPVT raw scores were transformed into standard equivalents (*SSEs*), a measure of deviation from the mean of the norming group (*M* = 100). Average *SSEs* were 103 for NH participants and 94 for HI participants [*t*(14) = -1.09, *p* = .292.]

In addition, HI participants were screened to have (1) bilateral profound congenital sensorineural hearing impairment (greater than 90 dB HL pure tone average across 500, 1000, and 2000 Hz) and (2) education in a mainstream and/or oral program for 8 or more years.

NH participants. Eight NH participants were recruited from among undergraduate students at California State University, Northridge (CSUN). The group mean age was 23.5 years (range, 21–28).

HI participants. Eight HI participants were recruited from among undergraduate students at CSUN. The group mean age was 22.5 years (range, 19–26). Reported age of hearing impairment onset was birth (5 participants), 1–2 years (2 participants), exact age unknown but believed to be at birth (1 participant). Six participants had 100 dB HL or greater pure tone averages in both ears. One participant had a pure tone average of 98 dB HL in the left ear and of greater than 100 dB HL in the right ear, whereas another participant had a pure tone average of 95 dB HL in the right ear and of greater than 100 dB HL in the left ear. Thus, these participants had profound hearing impairments and relied primarily on vision for speech communication.

Stimuli

The stimuli were monosyllabic and initial-stress disyllabic words chosen from the 35,000-word PhLex database (Seitz, Bernstein, Auer, & MacEachern, 1998), contrasted on their frequency of occurrence in English (low frequency vs. high frequency) and on the number of words in their LEC (unique, medium, and large LEC sizes). All words had a Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984) familiarity score greater than 5 (on a scale from 1 [*unknown word*] to 7 [*very familiar word*]). The number of words per cell, together with their average frequency (Kučera & Francis, 1967), is reported in Table 1 (see the Appendix for an entire list of stimuli). LEC size was computed as the number of words in PhLex that were predicted to be visually highly similar to the target word, as will be described below.

Lexical confusability was estimated using the computational method in Auer and Bernstein (1997). The method involves three steps. (1) Rules are developed to retranscribe words so that their transcriptions represent only the segmental distinctions that are estimated to be visually perceivable. The retranscription rules are in the form of PECs. (2) Retranscription rules are applied to the words in a phonemically transcribed, computer-readable lexicon. (3) The retranscribed words are sorted so that words rendered identical (no longer distinct) are placed in the same LEC.

Transcription rules (Step 1) were generated from a very large database of phoneme confusion data (Auer, Bernstein, Waldstein, & Tucker, 1997) for the talker who spoke the words of the present study. The consonants' perceptual confusions were obtained from disyllabic nonsense stimuli lipreading responses, and the vowels' were obtained from monosyllabic nonsense stimuli lipreading responses. Vowel stimuli included /i, ɪ, e, æ, a, ʌ, ɔ, ə, o, u, ə/ and the rhotized vowels /ɪr, ar, ær, ur, ɔr, ʊr, aʊr, aɪr, er, ər/. Rhotized vowels were part of the set, because of their potential to produce different patterns of similarity than do their nonrhotized equivalents. Consonant stimuli included /b, p, m, f, v, r, θ, ð, ʃ, d, l, n, s, z, t, j, w, h, g, k/. Included also were consonant clusters selected to

Table 1
Average Lexical Frequency (LF) of the Stimuli per Lexical Equivalence Class Size (Unique, Medium, Large) for Monosyllables and Disyllables (With Number of Stimuli)

Frequency	Monosyllables						Disyllables					
	Unique		Medium		Large		Unique		Medium		Large	
	LF	No.	LF	No.	LF	No.	LF	No.	LF	No.	LF	No.
High	195	25	194	25	198	25	196	25	195	25	69	16
Low	6	25	6	25	6	25	6	25	5	25	7	16

Table 2
Phoneme Equivalence Classes (PECs) Used
in the Retranscription of the PhLex Database

Vowel PECs

{i, ɪ}, {e, æ}, {ɑ, ʌ}, {ɔ, ɑr, ɔʊ}, {ur, u, o}, {u}, {ɛr}, {ɪr}, {er},
 {eɪ}, {aɪ}, {ɔɪ}

Consonant PECs

PECs for Word-Initial Consonants

If initial vowel was one of /u, o, ur, ɛr, u, ɔr, ər/:

{b, p, m, pr}, {f, v, r}, {θ, ð}, {tʃ, dʒ, ʃ, d, l, n, s, z, t, j, st, tr},
 {w, h, g, k, gr, kr}

If initial vowel was one of /æ, ɑ, ʌ, ɔ, ɑr, aɪ, ɔɪ, ɔʊ, ɔr, aɪr, ər/:

{b, p, m}, {f, v}, {θ, ð}, {tʃ, dʒ, ʃ}, {w}, {r, gr, k},
 {d, n, s, g, k, z, t, j, st, tr}, {h}, {l}, {pr}

If initial vowel was one of /i, ɪ, e, ɪr, er, eɪ/:

{b, p, m}, {f, v}, {tʃ, dʒ, ʃ}, {θ, ð}, {w, r}, {gr, kr},
 {d, h, n, s, g, k, z, t, j, st, tr}, {l}, {pr}

If initial vowel was /ə/:

{b, p, m, pr}, {f, v}, {θ, ð, l},
 {tʃ, dʒ, ʃ, w, r, gr, kr, d, h, n, s, g, k, z, t, j, st, tr}

PECs for Noninitial Consonants

{b, p, m}, {f, v}, {θ, ð}, {tʃ, dʒ, ʃ}, {w, r}, {d, n, s, g, k, z, t, j}

Note—Each pair of curly brackets defines a PEC.

represent the most frequent position-specific clusters in English, based on token counts in the Brown corpus (Kučera & Francis, 1967). The consonant clusters were /pr, st, tr, gr, kr, nd, ns, nt, kt, ls/. The PECs were derived by the technique of hierarchical cluster analysis (Aldenderfer & Blashfield, 1984). The level of clustering selected for grouping phonemes into PECs was such that all PECs included 75% or more within-cluster responses. Word-initial consonant PECs were generated in a manner that took the identity of the initial vowel (i.e., the vowel in the first syllable) into account. That is, monosyllable and disyllable stimuli were generated while taking into account the effects of vowel context on prevocalic consonant confusions. PECs are reported in Table 2.

In Step 2, all of the monosyllabic and disyllabic words in PhLex (Seitz et al., 1998) were retranscribed according to the PECs in Table 2. In Step 3, the test words were organized into three LEC-size categories, based on their retranscribed format. *Unique-LEC* words were from lexical equivalence classes of size one—that is, they did not have any visual competitors, as defined by the above phoneme clustering. *Medium-LEC* words were from LECs of size 2–6. *Large-LEC* words were from LECs of size 10–60. Because of the limited number of available disyllabic words in large LECs, only 32 stimuli were used in the disyllabic low-frequency (16) and high-frequency (16) large-LEC categories (see Table 1). This limitation influenced the overall frequency of the stimuli in these categories, with a resulting lower mean frequency for the high-frequency stimuli of the large-LEC disyllable category than for the high-frequency stimuli in the two other categories. Another consequence was that the mean LEC size of the large-LEC words was 34.4 for the monosyllables and only 13.2 for the disyllables.

Procedure

All the participants were tested individually at CSUN in a quiet room. They were seated in front of a computer monitor and were given verbal instructions. A certified sign language interpreter or a deaf research assistant administered the instructions to the HI participants, using English signs in synchrony with speech. The 282 video-recorded words, presented one at a time, were spoken by a female talker, with her face filling most of the monitor frame. Words were

presented in four blocks. Two blocks contained the monosyllables (75 words in each), and the other two contained the disyllables (66 words in each). Proportions of high- versus low-frequency words, and unique-, medium-, and large-LEC words were identical across all blocks. Block presentation order was rotated across participants. Within each block, word presentation was randomized for each participant. The experiment began with a practice block of 10 monosyllables and one of 10 disyllables. For both the practice and the experimental blocks, the participants were asked to identify each word in an open-set format by typing it in on a computer keyboard. They were told that all of the stimuli were words and were therefore encouraged to provide a word response, but they were allowed to enter a nonword response if they could not perceive a word that corresponded to the input. After entering a response, the participants pressed a keyboard key to see the next word.

RESULTS AND DISCUSSION

Word Identification Scores

All the responses were screened by two people for misspellings or obvious typographical errors. These errors were corrected when both referees agreed that there was no ambiguity concerning the intended response. The responses were then coded as correct or incorrect. Incorrect responses included any departure from the target word, such as another word, a nonsense word, an untranscribable response (e.g., *wqxa*), or no response. The percentage of correct responses was calculated for each cell of the design, examining group (NH, HI), word LEC size (unique, medium, large), word frequency (high, low), and word length (monosyllable, disyllable). The results are reported in Table 3 and plotted in Figure 1.

Analyses of variance (ANOVAs) were performed on the identification scores by subjects (F_1) and by items (F_2). Analyses were also performed on the arcsine transformation of the identification scores in order to stabilize variance at the extremes of the proportions measures. The statistics on the latter are reported only if they notably departed from the analyses on the nontransformed identification scores.

Overall, words were identified more accurately when the LEC size was low and when the frequency of occurrence was high. Accordingly, ANOVAs revealed effects of LEC size [$F_1(2,28) = 239.45, F_2(2,270) = 63.50, p < .001$] and word frequency [$F_1(1,14) = 463.16, F_2(1,270) = 54.51, p < .001$]. There was a slight advantage (by items) for HI over NH participants [$F_1(1,14) < 1, F_2(1,270) = 3.64, p = .06$]. The group factor did not interact with any other factors. There was also an advantage for monosyllables over disyllables, significant by subjects [$F_1(1,14) = 13.14, p < .005$], but not by items [$F_2(1,270) = 2.33, p = .13$]. None of the factors interacted significantly, except for a LEC size \times word frequency interaction [$F_1(2,28) = 34.06, p < .001; F_2(2,270) = 4.07, p < .02$]. This interaction indicated that the frequency effect was less pronounced among words with a large-LEC size than in the two other LEC size categories [even though the frequency effect among large-LEC words was significant; $F_1(1,14) = 32.56, p < .001; F_2(1,96) = 8.66, p < .005$]. This inter-

Table 3
Mean Percentages and Ranges of Correct Word Identifications per Lexical Equivalence Class Size (Unique, Medium, Large)

Frequency	Monosyllables						Disyllables					
	Unique		Medium		Large		Unique		Medium		Large	
	<i>M</i>	Range	<i>M</i>	Range	<i>M</i>	Range	<i>M</i>	Range	<i>M</i>	Range	<i>M</i>	Range
Normal Hearing												
High	62	40–76	48	36–56	14	8–20	53	24–80	37	20–48	16	0–31
Low	38	20–48	19	4–36	6	0–12	33	16–56	16	4–24	2	0–6
Impaired Hearing												
High	62	48–72	47	20–68	19	4–48	58	48–76	47	32–64	11	0–25
Low	41	26–60	15	0–36	12	0–24	36	16–64	19	12–36	6	0–19

action probably results from the combination of the design-induced relatively low frequency of the high-frequency large-LEC disyllables (see Table 1) and a potential floor effect for the words in the low-frequency large-LEC category. The latter possibility is partly supported by the reduction in the *F* values of the LEC size \times word frequency interaction on the arcsine-transformed scores [$F_1(2,28) = 9.11, p < .002$; $F_2(2,270) = 2.78, p < .07$].

These results indicate that the difficulty with which visual spoken words are recognized cannot simply be attributed to a general reduction in intelligibility owing to visual similarity at the phoneme level. Indeed, performance was strongly influenced by the number of visually similar words in the lexicon and by the frequency of occurrence of the test words. Consistent with models that posit that word recognition is driven by a process of lexical discrimination (e.g., P. A. Luce & Pisoni, 1998), our results show that lipreading accuracy, too, is a function of perceptual similarity to word candidates in the mental lexicon. Word frequency exerted a considerable influence on recognition accuracy, even in the case of maximal lexical distinctiveness (LEC size of 1). This latter condition is most similar to auditory word recognition, in which words under good listening conditions are typically intelligible enough to be mutually distinctive. Thus, the widely documented word frequency effect occurs irrespective of the input modality and across varying levels of lexical distinctiveness in the test words.

Phoneme Identification Scores

To examine identification accuracy further, each response was coded in terms of percentage of phonemes correct. Phonemes-correct scores provide a more sensitive measure of speech perception, because they include information about phonetic processing performance on incorrect responses, which were all coded identically in the words-correct analyses. To calculate phoneme identification scores, the phonemic transcriptions of the stimuli and the responses (whether these were words, word fragments, or nonwords) were submitted to a software sequence comparison program (Bernstein, Demorest, & Eberhardt, 1994) that aligned each stimulus–response pair phoneme by phoneme. Sequence comparison takes into account differ-

ences in symbol strings owing to substitutions, deletions, and insertions. The software includes a minimization algorithm (Bernstein et al., 1994; Sankoff & Kruskal, 1983) that seeks the lowest total cost for aligning the phonemes from the stimulus and the response. For these scores, costs for insertions and deletions were selected so that only exact phoneme-to-phoneme alignments would occur.

A measure of percentage of phonemes correct was calculated for each response for each participant. Percentage of phonemes correct was the mean of the total correct phonemes in each response divided by the number of phonemes in the respective stimulus word. As in the word identification analyses, the mean percentage of phonemes correct was calculated as a function of group (NH, HI), word LEC size (unique, medium, large), word frequency (high, low), and word length (monosyllable, disyllable). The results are reported in Table 4 and plotted in Figure 2.

As before, ANOVAs were carried out both on the scores and on their arcsine transformation. Statistics on the latter are reported only if they depart from those on the former. The ANOVAs revealed patterns of results similar to those obtained for word identification accuracy, although HI participants differed from NH participants. HI individuals perceived phonemes more accurately than NH ones did [63% vs. 57%, respectively; $F_1(1,14) = 4.29, p = .057$; $F_2(1,270) = 50.70, p < .001$]. Consistent with the words-correct analyses, accuracy was greater when LEC size was low, when frequency was high, and when the stimuli were monosyllables [LEC size, $F_1(2,28) = 393.57, F_2(2,270) = 49.97, p < .001$; frequency, $F_1(1,14) = 78.03, F_2(1,270) = 22.31, p < .001$; word length, $F_1(1,14) = 45.44, p < .001$, and $F_2(1,270) = 10.43, p < .002$]. LEC size and frequency were found to interact, but in the subjects analysis only [$F_1(2,28) = 9.89, p < .002$; $F_2(2,270) = 1.22, p = .30$]. As was suggested earlier, this interaction probably reflects the lower frequency of the disyllables in the large-LEC condition. None of the other interactions was significant (at the $p = .05$ level) in either subjects or items analyses.

This set of analyses underscores two facts. First, it is clear that the amount of phonetic information perceived in visual speech can be quite high. In particular, we found that frequent monosyllabic unique-LEC words generated

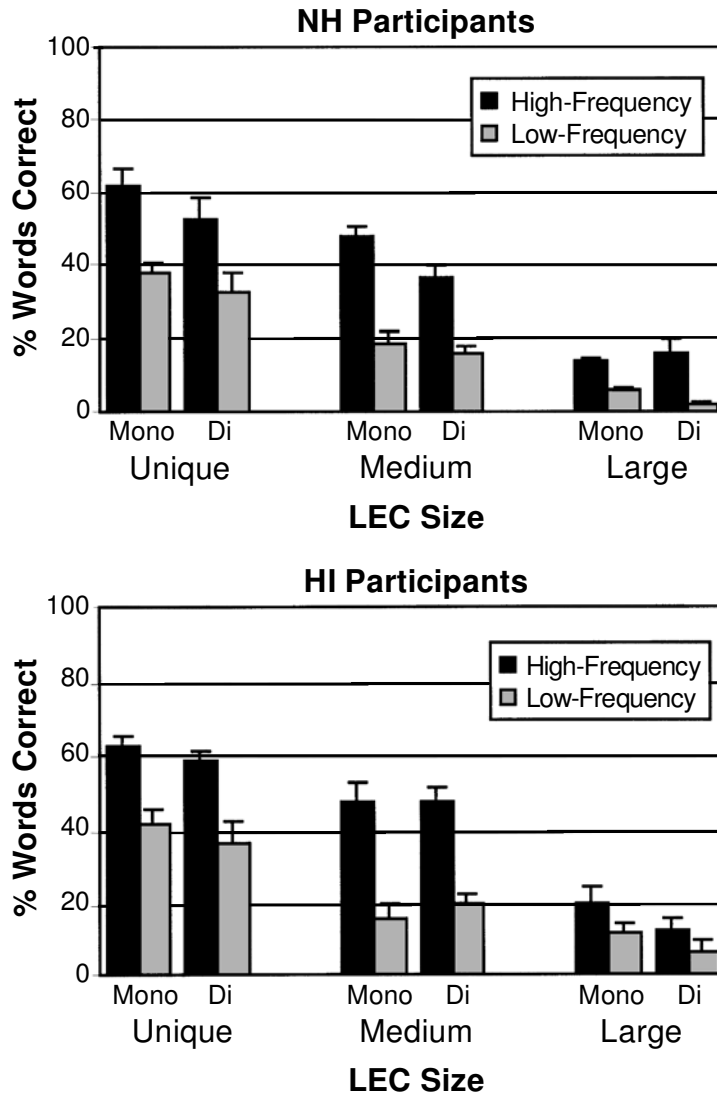


Figure 1. Mean percentages of words correct by subjects (and error bars) as a function of the lexical equivalence class (LEC) size, length, and frequency of occurrence of the test words. Upper panel, normal-hearing (NH) participants; lower panel, hearing-impaired (HI) participants. Mono, monosyllabic; Di, disyllabic.

an average of 82% phonemes correct among HI participants, with a maximum of 93% for 1 individual. Even in the low-frequency large-LEC condition, in which word identification scores were extremely low (about 6%; see Figure 1), analyses at the level of phonemes indicated that a considerable amount of phonetic information was successfully perceived (about 44% phonemes correct; see Figure 2). Thus, despite the phonetic impoverishment of visual speech, impressive levels of phonetic processing can be achieved. Moreover, these results corroborate the patterns of word recognition accuracy described above in showing that speech processing is contingent on lexical attributes such as stimulus-based similarity and word frequency of occurrence.

Second, the phonemes-correct analyses suggest a performance gap between HI and NH participants, which was noted previously for the words-correct analyses. Such a difference is consistent with the hypothesis that the necessity for deaf individuals to attend to visual information can result in enhanced visual phonetic perception (Bernstein, Demorest, & Tucker, 2000; Demorest & Bernstein, 1992) and is at odds with the competing assertion that normal hearing is necessary for achieving the highest levels of lipreading accuracy (Conrad, 1977; Mogford, 1987; Pelson & Prather, 1974). However, group differences here were inconsistent, suggesting that participant prescreening was successful in producing an NH group that was quite competent at the recognition task.

Table 4
Mean Percentages and Ranges of Phonemes Correct per Lexical Equivalence
Class Size (Unique, Medium, Large)

Frequency	Monosyllables						Disyllables					
	Unique		Medium		Large		Unique		Medium		Large	
	M	Range	M	Range	M	Range	M	Range	M	Range	M	Range
Normal Hearing												
High	78	66–87	70	57–79	46	36–57	73	56–92	59	47–66	42	20–55
Low	70	60–76	58	40–67	43	36–54	56	37–76	52	44–63	33	22–44
Impaired Hearing												
High	82	68–93	72	53–87	55	43–77	76	69–87	73	58–84	51	43–64
Low	74	63–84	58	46–70	53	39–63	64	48–82	57	50–72	45	35–61

Lexical Distinctiveness and Phonemic Intelligibility

The conclusion that the patterns of results were driven by lexical variables (the lexical distinctiveness hypothesis) needs to be verified against the competing hypothesis that overall phonetic intelligibility of the phonemes in test words can account for the results. For example, words from LECs of size one could be identified more accurately than words from larger LECs, because the former could turn out to be composed of visually more discriminable phonemes, independent of lexical distinctiveness. Should this be the case, there would be no need to invoke lexical competition to account for the above results.

An index of phonemic intelligibility was generated for each word by computing its mean phoneme equivalence class size $[(\sum \text{PECsize}_i)/n]$, where i indexes each phoneme, and n is the total number of phonemes in the word]. For this computation, we used the following set of PECs²: {d, n, s, g, k, z, t, j, st, tr}, {p, b, m}, {θ, ð}, {tʃ, dʒ, ʃ, ʒ}, {f, v}, {l}, {ŋ}, {h}, {w, r}, {ɔ, əʊ, ar}, {ur, ʊ, o}, {u}, {ɪr}, {ɜr}, {i, ɪ}, {ɜ, æ}, {ɑ, ʌ}. As an example, the four phonemes making up the word *film* are visually similar to two, two, one, and three other phonemes, respectively. Thus, the mean PEC size for *film* is 2.0. PEC values can be independent of lexical distinctiveness—LEC values. For instance, although *long* and *school* are both visually unique (LEC size = 1), *long* has a mean PEC size of 1.67, whereas *school* has a PEC size of 5.50. Therefore, *long* is, in principle, more visually intelligible than *school*, despite the fact that both words are visually distinct from all other words in the lexicon. For the following analyses, actual LEC sizes were used instead of the category names used earlier (unique-, medium-, and large-LEC). Likewise, frequencies were converted from “high” and “low” to the logarithmic value of their absolute frequency of occurrence in Kučera and Francis (1967).

The numeric values of PEC mean, LEC size, frequency, and word length were entered in several correlation analyses to assess the effect of each factor on item-level word recognition (w) and phoneme identification (p) scores. As was expected on the basis of the above ANOVA, both accuracy indices (stimulus means pooled across NH and HI participants) correlated highly with LEC size ($r_w = -.416, r_p = -.410, p < .001$) and frequency ($r_w = .359,$

$r_p = .244, p < .001$), but not reliably with word length ($r_w = -.025, \text{n.s.}; r_p = -.115, p = .053$). Accuracy was higher on words with fewer lexical competitors and with higher frequency of occurrence. However, accuracy also correlated with PEC mean ($r_w = -.421, p < .001; r_p = -.457, p < .001$), indicating that words and phonemes were recognized more accurately if the mean phonemic intelligibility of the words was high (i.e., low PEC mean). Although this correlation was expected in the case of phoneme identification accuracy, the correlation with word identification accuracy could suggest that word recognition was driven by phonemic intelligibility. A significant correlation between the PEC mean and LEC size of the words of our sample ($r = .590, p < .001$) substantiated the need for additional analyses.

In an attempt to isolate the effect of LEC size on identification accuracy independent of phonemic intelligibility, we calculated a partial correlation between LEC size and response accuracy, statistically controlling for PEC mean and word frequency. The correlation, which proved significant for both accuracy measurements ($r_w = -.238, p < .001; r_p = -.195, p < .002$), confirmed that lexical distinctiveness alone is reliably related to visual spoken word recognition and phoneme identification accuracy, independently of overall phonemic intelligibility and word frequency.

A similar analysis examining the correlation between word frequency and accuracy, controlling for LEC size and PEC mean, revealed that word frequency per se, too, correlated with recognition accuracy ($r_w = .402, p < .001; r_p = .277, p < .001$). Finally, mean PEC size itself, with LEC size and frequency controlled, correlated with accuracy ($r_w = -.266, p < .001; r_p = -.308, p < .001$). Thus, although phonemic intelligibility does affect recognition performance, lexical distinctiveness and word frequency clearly provide their own contributions to visual spoken word recognition over and above inherent phonemic confusion in the input.

Analysis of the Errors

A correlate to the hypothesis that visual spoken words are recognized through a process of lexical discrimination is that incorrect identification responses should fall within the predicted LEC of the target word more often than

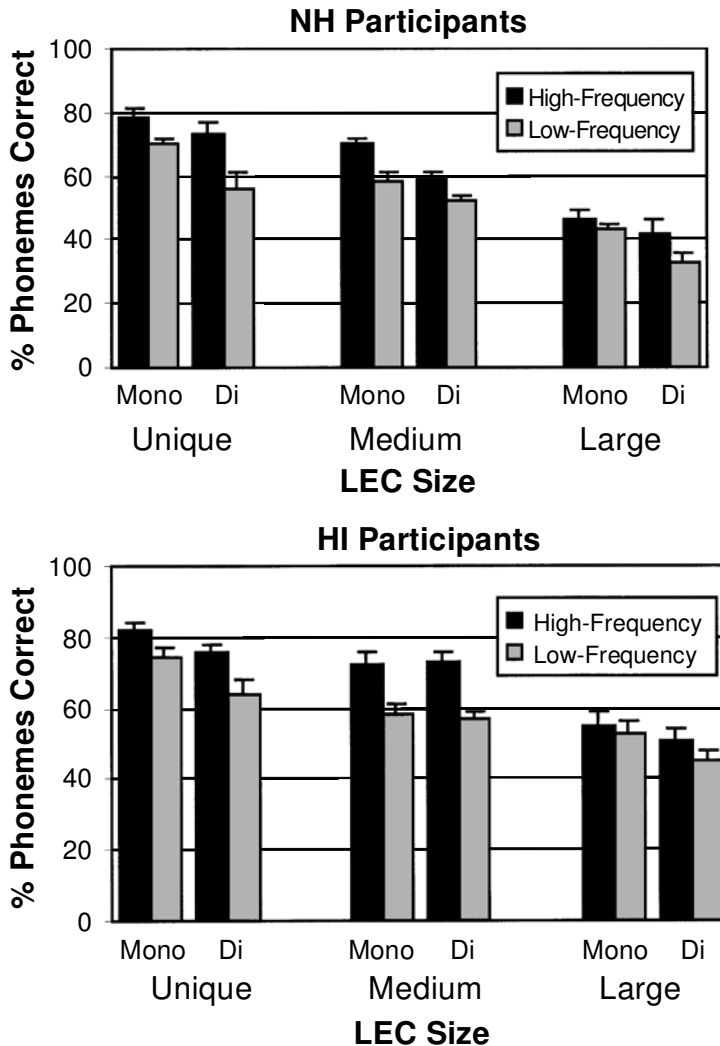


Figure 2. Mean percentages of phonemes correct by subjects (and error bars) across all word and nonword responses, as a function of the lexical equivalence class (LEC) size, length, and frequency of occurrence of the test words. Upper panel, normal-hearing (NH) participants; lower panel, hearing-impaired (HI) participants. Mono, monosyllabic; Di, disyllabic.

would be expected by chance. When incorrect responses to medium-LEC and large-LEC words were considered in this respect, the mean proportion of within-LEC error responses was .27 for monosyllables and .14 for disyllables (see Table 5). The chance level for each category, computed as the mean ratio between the number of words within the LEC of a target word and all the words of the same length (and same stress pattern) in the lexicon, was smaller than .00005 in both cases. Thus, when the participants failed to identify a word, they chose a word within the LEC of the target word far more often than would have been expected by chance, which further illustrates the constraining influence of stimulus-based lexical similarity on word recognition.

An ANOVA was performed on the within-LEC percentages of incorrect responses, examining group (NH,

HI), word LEC size (medium, large), word frequency (high, low), and word length (monosyllable, disyllable). The proportion of within-LEC errors was higher among HI than among NH participants [23.94% vs. 16.75%, respectively; $F_1(1,14) = 4.47, p = .05$; $F_2(1,174) = 12.94, p < .001$] and, as was mentioned above, was higher for monosyllables than for disyllables [$F_1(1,14) = 44.38, p < .001$; $F_2(1,174) = 11.84, p < .002$]. Neither the LEC-size factor nor any of the interactions reached significance by either subjects or items. The group effect suggests that HI individuals constructed a more accurate phonetic representation of the visual input, which allowed them frequently to generate a response visually compatible with the target word. The word length effect probably reflects the difference in LEC sizes in the large-LEC monosyllables (34.4) versus the large-LEC disyllables (13.2), which

Table 5
Mean Proportions of Errors Within Lexical Equivalence Class
(With Standard Deviations) by Participants

Frequency	Monosyllables				Disyllables			
	Medium		Large		Medium		Large	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Normal Hearing								
High	.31	.17	.23	.14	.06	.06	.10	.09
Low	.19	.11	.22	.07	.09	.06	.13	.10
Impaired Hearing								
High	.27	.14	.37	.22	.18	.16	.14	.06
Low	.27	.08	.30	.10	.17	.07	.22	.07

was constrained by the small sample of available large-LEC disyllables.

GENERAL DISCUSSION

Early investigations of spoken word recognition (e.g., Forster, 1976; Miller & Johnson-Laird, 1976) mostly emphasized the internal characteristics of words as the determining factors in modeling listeners' response patterns (e.g., word frequency, length, syntactic category, semantic information). However, it is now generally accepted that, in conjunction with the influence of their internal characteristics, words compete directly against each other for recognition in the mental lexicon. In this view, recognition is the process by which words in the mental lexicon are isolated as a function of their form-based similarity to other stimulus words (e.g., P. A. Luce, 1986; P. A. Luce & Pisoni, 1998; McClelland & Elman, 1986).

Although several issues arise as to which words compete and what the basis for competition is (Bard & Shillcock, 1993), competition is typically construed to be in relation to auditorily experienced form-based similarity. The results of the present experiment provide evidence that lexical similarity (LEC size), when estimated visually, is a determining factor for word recognition in the visual spoken modality. Analogous to auditory spoken words, visual spoken words are recognized in the context of words that are perceptually similar to them: Words from small LECs were recognized much more accurately than were words from larger LECs. Importantly, this result held even when intrinsic similarity at the phonemic level was factored out of the analyses. Thus lipreading performance cannot be accounted for solely by bottom-up phonemic similarity. Instead, our results are compatible with the lexical distinctiveness hypothesis, in which word candidates compete in the interplay between the format of the input representations and the content of the lexicon. That is, competition is not fixed; it is a dynamic phenomenon that interacts with the stimulus condition. For example, it is unlikely that word forms that are confusable in the spoken visual modality (e.g., *bat* and *pat*) will generate comparable confusion patterns when their auditory counterparts are played in a degraded environment. Iversen et al. (1998) found that the PECs pertaining to the visual modality are dramatically different from those ob-

tained from auditory speech delivered through a vocoder (see also Grant & Walden, 1996). That is, there are substantial variations in the membership of phonemes to PECs and, consequently, of words to LECs, because the patterns of phonemic confusion specific to each modality define distinct competition spaces.

In a different approach to estimating lexical and phoneme-level effects for visual spoken word recognition, Auer (in press) computed an index of visual distinctiveness, called *neighborhood density* (P. A. Luce, 1986), for a set of monosyllabic visual spoken words. Similar to P. A. Luce's (1986; P. A. Luce & Pisoni, 1998) work, neighborhood density estimates were generated from the application of NAM's choice rule (R. D. Luce, 1959) to phonemic transcriptions of the test words. Visual phonetic similarity was established via visual nonsense syllable perceptual confusion matrices. Twelve participants with profound hearing impairments and 12 participants with normal hearing identified sparse- and dense-neighborhood, isolated spoken words presented visually. The results revealed a correlation between lexical neighborhood density and visual spoken word identification accuracy in both deaf and hearing respondents. Sparse-neighborhood words (high distinctiveness) were identified more accurately than dense-neighborhood words (low distinctiveness). Importantly, the amount of variance accounted for by the visual neighborhood estimates was comparable to that reported previously for auditory spoken word recognition (P. A. Luce & Pisoni, 1998). However, when the visual neighborhood estimates were replaced with ones derived from perceptual confusion data from auditory speech perception in noise (from P. A. Luce, 1986), the correlation between neighborhood density and speechreading accuracy was dramatically reduced. Auer's results reinforce the conclusion that word recognition is achieved relationally, with words competing against each other as a function of their *form-based* similarity. Our results are consistent in relation to the concept of LEC (or predicted high similarity), a more direct index of the perceptual confusion embodied in visual speech (Auer & Bernstein, 1997). In addition, the present data show that recognition through form-based similarity generalizes to disyllabic stimuli as well.

Form-based lexical similarity was not the only factor that influenced visual spoken word recognition accuracy in the present study. The frequency effect indicates that,

similar to printed and auditory words (e.g., Balota & Chumbley, 1984; Broadbent, 1967; Forster, 1976; Glanzer & Ehrenreich, 1979; Gordon, 1983; Howes & Solomon, 1951; Savin, 1963; Soloman & Postman, 1952), lip-read words benefit from occurring frequently in the ambient language. Whether the frequency effect arises from inherent differences in activation thresholds among word units (e.g., Marslen-Wilson, 1987; Morton, 1969) or from biases on the decision process (e.g., Grosjean & Itzler, 1984; P. A. Luce & Pisoni, 1998) is not yet resolved. However, an important finding is that the frequency effect remained strong even in the case of visually unique target words. Because such words have theoretically no direct competitors that frequency could favor, the processing advantage for high-frequency words probably results from a word-internal bias, which operates independently of the lexical similarity effect discussed above.

Similarly, the frequency effect found in the medium and large LEC size conditions is comparable to frequency effects found for speech recognition in noise. In speech recognition in noise, ambiguous phonetic information causes a larger set of compatible word candidates to be available for selection, with word frequency being a primary selection factor (e.g., Broadbent, 1967). The frequency effect in the medium- and large-LEC words of our experiment could have resulted from a similar bias for choosing a high-frequency candidate among the members of a test word's LEC (Auer & Bernstein, 1997). Obviously, a prerequisite for this hypothesis is that the frequency distributions among the competitors (i.e., the other within-LEC words) of the high- and low-frequency test words should be equal. A computation of the frequency of the words within the LECs of the test words—in the medium- and large-LEC conditions—revealed that it was indeed the case: The mean log-transformed frequency of the within-LEC candidates was 29.34 (range, 1.67–90.20) in the high-frequency condition and 26.77 (range, 1.67–90.20) in the low-frequency condition. The mean log-transformed frequency of the high-frequency test words themselves was 50.43, and the mean log-transformed frequency of the low-frequency test words was 19.10, which are, respectively, higher and lower than their within-LEC competitors. Hence, the observed advantage for high-frequency words may originate from the tendency to choose a visually compatible word with relatively high frequency.

The frequency effect is also reflected in some of the incorrect responses. For example, the average frequency of incorrect responses falling within the LEC of a low-frequency target (e.g., responding *dog* instead of *stark*) was higher than that of the target itself ($M = 31.58$; range, 1.68–85.53). However, the corresponding figure in the high-frequency condition was lower than that of the targets ($M = 37.87$; range, 1.68–72.77), suggesting that, in some cases at least, factors other than frequency influenced the selection process among the within-LEC candidates. Previous research (Bernstein, Iverson, & Auer, 1997) indicates that one such factor could be that

lipreaders perceive finer visual phonetic cues (e.g., subphonemic and coarticulatory cues) than those used to establish the LECs. As a result, a test word might actually have fewer lexical competitors than was estimated by the LEC statistic, hence affecting the frequency distribution of the candidates available for recognition.

The present experiment demonstrates not only that lexical similarity and word frequency affect word recognition beyond the auditory modality, but also that these factors operate regardless of the participant's long-term experience with the auditory modality. Respondents with normal and impaired hearing showed comparable sensitivity to LEC size and word frequency. This result indicates that the visual discrimination strategies used by NH individuals when lip-reading words are not altered by the patterns of similarity they encounter in the ambient auditory environment. Both groups derived their responses from the test words' visual similarity with the other words of the lexicon. Thus, lexical competition has to be defined with respect to the modality of entry during processing, not to the modality in which words were learned.

The fact that lexical competition takes place at the crossroad between the lexicon and the input signal has consequences for models of word recognition aiming to model lipreading. First, our findings show that it is necessary to define clearly what phonetic information is available in the signal—and how it is distributed in the signal—in order to predict the patterns of similarity between word candidates. Obviously, this requirement applies to both visual and auditory speech. Models of auditory spoken word recognition typically use phonemes as a basis to evaluate the fit between the input and the memory representations (P. A. Luce & Pisoni, 1998; Marslen-Wilson & Welsh, 1978), but subphonemic units have also been proposed (e.g., Klatt, 1980; Marslen-Wilson, 1987, 1993; McClelland & Elman, 1986; Norris, 1994). The few studies that have investigated lip-read word recognition, including the present one, have used groupings of mutually confusable phonemes, sometimes referred to as *visemes* (Fisher, 1968; Massaro, 1998) or, more generally, PECs, to compute lexical similarity. Although such approaches to lipreading have proved informative, they probably remain a coarse approximation of the perceptual experience involved in lipreading. For example, Bernstein et al. (1997) reported that lipreaders show some sensitivity to subviseme information (e.g., the distinction between *bite* and *mite*), which suggests that the viseme construct may underestimate the phonetic information available in visual speech. More accurate estimates of the phonetic information available to perceivers will help better predict lexical similarity and, hence, the extent of lexical competition. Second, the fact that word lipreading was found to be driven by the same patterns of visual similarity in HI and NH individuals is consistent with the idea that competition does not involve lexical representations organized according to the attributes of a single modality. Instead, our results show that form-based lexical distinctiveness constitutes a valid word recognition mechanism regardless of

(1) the perceptual modality, (2) the specific patterns of phonetic similarity, and (3) the hearing status of the perceiver (deaf vs. hearing).

REFERENCES

- ALDENDERFER, M. S., & BLASHFIELD, R. K. (1984). *Cluster analysis*. Beverly Hills, CA: Sage.
- AUER, E. T., JR. (in press). The influence of the lexicon on speechread word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*.
- 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.
- AUER, E. T., JR., BERNSTEIN, L. E., WALDSTEIN, R. S., & TUCKER, P. E. (1997). Effects of phonetic variation and the structure of the lexicon on the uniqueness of words. In C. Benoît & R. Campbell (Eds.), *Proceedings of the ESCA/ESCAP workshop on audio-visual speech processing* (pp. 21-24). Rhodes, Greece.
- BALOTA, D. A., & CHUMBLEY, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.
- BARD, E. G., & SHILLCOCK, R. C. (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G. T. M. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 235-275). Hove, U.K.: Erlbaum.
- BERGER, K. W. (1972). Visemes and homophenous words. *Teacher of the Deaf*, **70**, 396-399.
- 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.
- 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: The psychology of speechreading and auditory-visual speech* (pp. 211-228). Hove, U.K.: Psychology Press.
- BERNSTEIN, L. E., DEMOREST, M. E., & TUCKER, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, **62**, 233-252.
- BERNSTEIN, L. E., & EBERHARDT, S. P. (1986). *Johns Hopkins Lipreading Corpus I-II: Disc 1* [Videodisc]. Baltimore: Johns Hopkins University.
- BERNSTEIN, L. E., IVERSON, P., & AUER, E. T., JR. (1997). Elucidating the complex relationships between phonetic perception and word recognition in audiovisual speech perception. In C. Benoît & R. Campbell (Eds.), *Proceedings of the ESCA/ESCAP workshop on audio-visual speech processing* (pp. 21-24). Rhodes, Greece.
- BREEUWER, M., & PLOMP, R. (1985). Speechreading supplemented with formant-frequency information from voiced speech. *Journal of the Acoustical Society of America*, **77**, 314-317.
- BROADBENT, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, **74**, 1-15.
- CLUFF, M. S., & LUCE, P. A. (1990). Similarity neighborhoods of spoken two syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 551-563.
- CONRAD, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology*, **47**, 60-65.
- DAVIS, H., & SILVERMAN, S. R. (1970). *Hearing and deafness*. New York: Holt, Rinehart & Winston.
- DEMOREST, M. E., & BERNSTEIN, L. E. (1992). Sources of variability in speechreading sentences: A generalizability analysis. *Journal of Speech & Hearing Research*, **35**, 876-891.
- DUNN, L. M., & DUNN, L. M. (1981). *Peabody picture vocabulary test-revised*. Circle Pines, MN: American Guidance Service.
- FISHER, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech & Hearing Research*, **11**, 796-804.
- FORSTER, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North Holland.
- FORSTER, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-85). Hillsdale, NJ: Erlbaum.
- GLANZER, M., & EHRENREICH, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **18**, 381-398.
- GORDON, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning & Verbal Behavior*, **22**, 24-44.
- GRANT, K. W., & BRAIDA, L. D. (1991). Evaluating the articulation index for auditory-visual input. *Journal of the Acoustical Society of America*, **89**, 2952-2960.
- GRANT, K. W., & WALDEN, B. E. (1996). Spectral distribution of prosodic information. *Journal of Speech & Hearing Research*, **39**, 228-238.
- GROSJEAN, F., & ITZLER, J. (1984). Can semantic constraint reduce the role of word frequency during spoken-word recognition? *Bulletin of the Psychonomic Society*, **22**, 180-182.
- HOWES, D. H. (1954). On the interpretation of word frequency as a variable affecting speech recognition. *Journal of Experimental Psychology*, **48**, 106-112.
- 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.
- HOWES, D. H., & SOLOMAN, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, **41**, 401-410.
- IVERSON, P., BERNSTEIN, L. E., & AUER, E. T., JR. (1998). Phonetic perception and word recognition. *Speech Communication*, **26**, 45-63.
- JACKSON, P. L. (1988). The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *Volta Review*, **90**, 99-115.
- KLATT, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.
- KRICOS, P. B., & LESNER, S. A. (1982). Differences in visual intelligibility across talkers. *Volta Review*, **84**, 219-225.
- KRICOS, P. B., & LESNER, S. A. (1985). Effect of talker differences on the speechreading of hearing-impaired teenagers. *Volta Review*, **87**, 5-16.
- KUCERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LUCE, P. A. (1986). *Neighborhoods of words in the mental lexicon* (Research on Speech Perception, Tech. Rep. No. 6). Bloomington: Indiana University, Department of Psychology, Speech Research Laboratory.
- 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.
- LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- MARSLÉN-WILSON, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71-102.
- MARSLÉN-WILSON, W. D. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187-210). Hove, U.K.: Erlbaum.
- MARSLÉN-WILSON, W. D., & WARREN, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675.
- MARSLÉN-WILSON, W. D., & WELSH, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.
- MASSARO, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MCQUEEN, J. M., CUTLER, A., BRISCOE, T., & NORRIS, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language & Cognitive Processes*, **10**, 309-331.
- MILLER, G. A., & JOHNSON-LAIRD, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

- 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: Erlbaum.
- 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.
- MORTON, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76**, 165-178.
- MORTON, J. (1979). Word recognition. In J. Morton & J. D. Marshall (Eds.), *Psycholinguistics: 2. Structures and processes* (pp. 107-156). Cambridge, MA: MIT Press.
- NITCHE, E. B. (1916). The use of homophenous words. *Volta Review*, **18**, 85-93.
- NORRIS, D. (1994). Shortlist: A connectionist model of continuous word recognition. *Cognition*, **52**, 189-234.
- 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 Rep. No. 10). Bloomington: Indiana University, Department of Psychology.
- PELSON, R. O., & PRATHER, W. F. (1974). Effects of visual message-related cues, age, and hearing impairment on speechreading performance. *Journal of Speech & Hearing Research*, **17**, 518-525.
- SANKOFF, D., & KRUSKAL, J. B. (1983). *Time wraps, string edits, and macromolecules: The theory and practice of sequence comparison*. Reading, MA: Addison-Wesley.
- SAVIN, H. B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, **35**, 200-206.
- SEITZ, P. F., BERNSTEIN, L. E., AUER, E. T., JR., & MACEACHERN, M. (1998). *PhLex (Phonologically Transformable Lexicon): A 35,000-word pronouncing American English lexicon on structural principles, with accompanying phonological rules and word frequencies*. Los Angeles: House Ear Institute.
- SOLOMAN, R. L., & POSTMAN, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, **43**, 195-201.
- WANG, M. D., & BILGER, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, **54**, 1248-1266.

NOTES

1. Wide individual differences in visual spoken word recognition are found in the hearing and deaf populations (Bernstein, Demorest, & Tucker, 2000). Random selection of participants would virtually guarantee that some individuals would perform at extremely low levels of accuracy. For the present experiment, we drew our NH and HI participant samples from corresponding segments in their respective parent populations to ensure comparable base performance in both groups, and because we sought to observe effects in individuals who could all competently perform the lipreading task.
2. This set of PECs is simpler than the one used to recode the lexicon and select the test words. These PECs do not reflect sensitivity to the identity of the initial vowel of the word. They were obtained from estimates of visual similarity among nonsense sequences produced by the same talker as the one used in the present experiment. This simpler set was chosen because a context-sensitive recoding algorithm can generate multiple PEC-size values for the same phoneme within a single word and make it difficult to choose the value that is most appropriate.

APPENDIX

List of Stimuli

Monosyllables

Unique LEC

High frequency: speech brief strength friend floor month strange fare hung far growth long form school film spring frame file both roof charge page smile square farm

Low frequency: swam lump plunge sprung crisp famed swamp froze rape clutch crouch breadth dwarf shrink thrill breathe hunch coil thrift sponge booth shrill twelfth sworn shriek

Medium LEC

High frequency: hoarse price fall stage march full point live staff class core drive serve space food voice care brown sure line late bill force give health

Low frequency: grudge strife wink comb rude spine bump lied probe browse mink clad pulp pierce ramp hanged brute strewn stealth lace sunk burnt punch breech robe

Large LEC

High frequency: hit site peace news stand mean sound keep meat bad note best tried sent gone son met tax needs soon shone gun stock dark case

Low frequency: gaunt tag bust truce dip mint putt soak hood stark teens tease bout tread bean stud stain mast stint goat tart peck hook tilt hid

Disyllables

Unique LEC

High frequency: normal famous foreign process trouble knowledge current woman private children special college moment social science student function problem southern central question spirit product thousand congress

Low frequency: vibrant chestnut cherish marvel straighten fragment bankrupt ruthless township garbage captive tribal zenith trumpet junction scaffold fortress grandson garment symptom trousers blizzard piping joyous mischief

APPENDIX (Continued)

Medium LEC

High frequency: human surface table series common nation music million simple certain district person western modern running present reading purpose morning husband service final single coming working

Low frequency: locust diction rattle graphic puppet vantage chuckle ripple shovel wallet tenant drizzle placid jagged widen siding swivel stricken tariff digit triple vanish fluent cunning tubing

Large LEC

High frequency: season beaten panic gotten market model pocket battle mussel hidden basis subtle basic senate dozen saddle

Low frequency: kitten mutton pedal bargain tackle tunnel satin basin buckle bucket bacon muzzle button deacon beacon menace

(Manuscript received December 20, 2000;
revision accepted for publication August 21, 2001.)