Speech perception without hearing

LYNNE E. BERNSTEIN
House Ear Institute, Los Angeles, California

MARILYN E. DEMOREST
University of Maryland, Baltimore County, Catonsville, Maryland

and

PAULA E. TUCKER
House Ear Institute, Los Angeles, California

In this study of visual phonetic speech perception without accompanying auditory speech stimuli, adults with normal hearing (NH; $n = 96$) and with severely to profoundly impaired hearing (IH; $n = 72$) identified consonant–vowel (CV) nonsense syllables and words in isolation and in sentences. The measures of phonetic perception were the proportion of phonemes correct and the proportion of transmitted feature information for CVs, the proportion of phonemes correct for words, and the proportion of phonemes correct and the amount of phoneme substitution entropy for sentences. The results demonstrated greater sensitivity to phonetic information in the IH group. Transmitted feature information was related to isolated word scores for the IH group, but not for the NH group. Phoneme errors in sentences were more systematic in the IH than in the NH group. Individual differences in phonetic perception for CVs were more highly associated with word and sentence performance for the IH than for the NH group. The results suggest that the necessity to perceive speech without hearing can be associated with enhanced visual phonetic perception in some individuals.

Viewing a talker is generally thought to afford extremely impoverished phonetic information, even though visual speech stimuli are known to influence heard speech (see, e.g., Dekle, Fowler, & Funnell, 1992; Green & Kuhl, 1989; MacLeod & Summerfield, 1990; McGurk & MacDonald, 1976; Middleweerd & Plomp, 1987; Rosenblum & Fowler, 1991; Sekiyama & Tohkura, 1991; Sumby & Pollack, 1954; Summerfield, 1987; Summerfield & McGrath, 1984). The literature on lipreading (speechreading) reports that visual speech stimuli are so impoverished that segmental distinctiveness is highly reduced (Fisher, 1968; Kuhl & Meltzoff, 1988; Massaro, 1998; Owens & Blazek, 1985), that accuracy for lipreading sentences rarely exceeds 10%–30% words correct (Rönnberg, 1995; Rönnberg, Samuelsson, & Lyxell, 1998), that relying on visual speech out of necessity, as a consequence of deafness, does not result in any substantial experientially based improvement in visual speech perception (Clouser, 1977; Conrad, 1977; Green, 1998; Lyxell & Rönnberg, 1991a, 1991b; Massaro, 1987; Mogford, 1987; Rönnberg, 1995; Rönnberg, Öhngren, & Nilsson, 1982; Summerfield, 1987; cf. Pelson & Prather, 1974; Tillberg, Rönnberg, Svärd, & Ahlner, 1996), and that, to achieve even moderate accuracy, lipreaders must rely on top-down psycholinguistic processes, nonlinguistic contexts, and strategic processes, such as guessing (Jeffers & Barley, 1971; Lyxell & Rönnberg, 1987; Rönnberg, 1995).

The present study questions this characterization—in particular, whether high levels of visual phonetic perception are impossible and whether examples of enhanced lipreading accuracy are not, in fact, associated with hearing impairment. Our skepticism about the characterizations in the literature arose in the course of studies on vibrotactile devices to aid lipreading, in which adults with congenital or early-onset profound hearing impairment were employed (Bernstein, Coulter, O'Connell, Eberhardt, & Demorest, 1993; Bernstein, Demorest, Coulter, & O'Connell, 1991). In formal testing and informal communication, these individuals demonstrated unexpectedly accurate lipreading, without the vibrotactile devices. In our research on lipreading in hearing adults (Demorest & Bernstein, 1992; Demorest, Bernstein, & DeHaven, 1996), the best lipreaders never outperformed the proficient deaf lipreaders in our vibrotactile aid studies. Scores for the most accurate deaf[1] participants in the unaided conditions were approximately 80% words correct in sentences (Bernstein et al., 1993; Bernstein et al., 1991).

The Impoverished Visual Phonetic Speech Stimulus

In a review of audiovisual speech perception, Kuhl and Meltzoff (1988) commented, "A relatively small proportion of the information in speech is visually available" (p. 240). Phoneme identification rates for nonsense syllable stimuli, depending on phonetic context, have been reported to be below 50% correct (e.g., rates range between 21% and 43% in Auer, Bernstein, Waldstein, & Tucker, 1997, and between 19% and 46% in Owens & Blazek, 1985; see, also, Benguerel & Pichora-Fuller, 1982; Fisher, 1968; Lesner & Kricos, 1981; Walden, Erdman, Montgomery, Schwartz, & Prosek, 1981; Wozniak & Jackson, 1979). The percentage for the correct identification of vowels in /h/V/g/ stimuli (where V = vowel) by adults with normal hearing, in a study by Montgomery and Jackson (1983), varied between 42% and 59% (see, also, Montgomery, Walden, & Prosek, 1987). However, Auer et al. (1997) reported 75% correct vowel identification for 19 vowels, including r-colored vowels across four phonetic contexts.

The term *viseme* was coined to refer to mutually confused phonemes that are deemed to form a single perceptual unit (Fisher, 1968; Massaro, 1998). When visemes are defined via cluster analyses by within-cluster response rates of 75% or greater, as was done by Walden, Prosek, Montgomery, Scherr, and Jones (1977), consonants have been found to fall into approximately six viseme clusters (i.e., /f v/, /θ ð/, /w r/, /p b m/, /ʃ ʒ tʒ tʃ/, /t d s z j k n g l/).

Whether phonemes within viseme groups are discriminable or not has not been examined as a general question. Nevertheless, that the viseme is a perceptual unit has been accepted in the literature. For example, Massaro (1998) asserts, "Because of the limited data available in visible speech as compared to audible speech, many phonemes are virtually indistinguishable by sight ... and so are expected to be easily confused. To eliminate these nonserious confusions from consideration, we group visually *indistinguishable* [emphasis added] phonemes into categories called *visemes*" (p. 394).

The viseme is a plausible perceptual unit, from the perspective of conventional accounts of articulatory phonetics: Many of the speech-related activities of the vocal tract are occluded from view by the lips, cheeks, and neck. Vocal fold vibration, for example, which is responsible for the voice fundamental frequency and which contributes to phonological distinctions between voiced and unvoiced consonants (e.g., /g/ versus /k/, /b/ versus /p/, and /z/ versus /s/; Lisker & Abramson, 1964), is completely invisible.[2] The type of vocal tract closure made by the tongue, which contributes to phonological manner distinctions (Catford, 1977), is only partially visible. For example, the acoustic distinction between /s/ and /t/ is due, in part, to the continuous maintenance of an air channel in the former case and the achievement of a brief closure in the latter, both at approximately the same tongue position in the vocal tract. It is difficult to see whether the tongue is completely blocking the air chan-

nel or closely approximating a blockage. Place of articulation of the tongue is also only partially visible. For example, tongue positions responsible for distinctions between consonants with dental versus velar constrictions (e.g., /t/ versus /k/) are difficult to view. Also, the state of the velum, which is responsible for nasality, is completely invisible. Thus, a major articulatory contributor to distinctions such as /m/ versus /b p/ is invisible in syllable-initial position (cf. Scheinberg, 1988).

The classical articulatory phonetic characterization does not, however, preclude other possible sources of visual phonetic information, such as the kinematics of the jaw, the cheeks, and the mouth. Indeed, Vatikiotis-Bateson, Munhall, Kasahara, Garcia, and Yehia (1996) and Yehia, Rubin, and Vatikiotis-Bateson (1997) have shown surprisingly high quantitative associations between external facial motions and speech acoustics. They reported, for example, that 77% of the variance in acoustic line spectrum parameters could be accounted for with orofacial movement data from the lips, jaw, and cheek. Consistent with the potential for observing visual phonetic detail, Bernstein, Iverson, and Auer (1997) showed that words predicted on the basis of viseme analyses to be visually indiscriminable were identified, on average, at approximately 75% correct (chance = 50%) in a two-interval forced-choice procedure. Some participants scored 100% correct. Thus, it is likely that the classical articulatory phonetics account of optical phonetic effects is not adequate and that other sources of phonetic information are potentially available to lipreaders.

Low Accuracy for Visual Perception of Words

There are many reports of inaccurate lipreading of connected speech (e.g., Breeuwer & Plomp, 1986; Demorest & Bernstein, 1992; Rabinowitz, Eddington, Delhorne, & Cuneo, 1992; Rönnberg, 1995; Rönnberg et al., 1998). Breeuwer and Plomp measured lipreading in a group of Dutch adults with normal hearing, using short sentences. The percentage of correct syllables was 10.6% on the first presentation of the sentences and 16.7% on the second. Demorest and Bernstein (1992) reported on 104 college students with normal hearing, who attempted to lipread words in sentences. Overall, the students identified 20.8% of the words. However, individual scores ranged between 0% and 45% words correct. Rabinowitz et al. (1992) studied lipreading in 20 adults with profound hearing impairments acquired after language acquisition. These adults had all received cochlear implants (devices that deliver direct electrical stimulation to the auditory nerves) and were tested on identification of words in sentences by vision alone in one of the conditions of the study. Two sets of materials were administered, differing in sentence length, vocabulary, and semantic properties. Performance was 18% words correct for the more difficult materials, with a range of scores from 0% to 45%, and a mean of 44% for the easier materials, with a range of 0% to 75%. Although the means

cited here support the generalization that lipreading is highly inaccurate, the ranges suggest that large, potentially functionally significant individual differences can occur.

Individual Differences

Individual differences in lipreading are not generally thought to be due to explicit training, although in the classical lipreading literature (Jeffers & Barley, 1971), it was assumed that training on speech patterns was required in order to master lipreading. That is, training was considered necessary to express a certain inborn potential. But because some people are apparently unable to benefit from training over years of experience (see, e.g., Heider & Heider, 1940), it is said that good lipreaders are born and not made.

That lipreading is an inborn capability is supported by reports that people whose hearing impairments are severe to profound are no more accurate in perceiving speech by eye than are people with normal hearing (Conrad, 1977; Lyxell & Rönnberg, 1989; Rönnberg, 1995; Summerfield, 1991). Perceivers with normal hearing are thought to be *more* accurate than perceivers with hearing impairments: "That auditory experience of speech enhances visual speech recognition is shown by the fact that the hearing are more competent at lip-reading [sic] than the deaf" (Mogford, 1987, p. 191). Given that visible speech influences hearing perceivers (see, e.g., Dekle et al., 1992; Green & Kuhl, 1989; MacLeod & Summerfield, 1990; Massaro, 1987; McGurk & MacDonald, 1976; Middleweerd & Plomp, 1987; Rosenblum & Fowler, 1991; Sekiyama & Tohkura, 1991; Sumby & Pollack, 1954; Summerfield, 1991; Summerfield & McGrath, 1984), if it afforded adequate phonetic information for speech perception, one might expect to observe its adequacy and, possibly, even its enhancement of performance among those deaf people forced to rely on it (J. L. Miller, 1991; Summerfield, 1991). Recently, Rönnberg (1995) reviewed the literature concerning this question and concluded that there was no evidence for enhanced performance in relation to auditory experience. Our experience and observations suggested that evidence could be obtained for enhancement associated with hearing impairment.

The Present Study

The emphasis in the present study was on phonetic perception. This contrasts with recent studies of lipreading by Rönnberg and colleagues (e.g., Lyxell & Rönnberg, 1987, 1992; Rönnberg, 1995), who sought to explain individual differences in lipreaders in terms of top-down psycholinguistic and strategic processes. According to their view, the visual stimulus affords so little phonetic information that performance can only be optimized at a later stage than perceptual processing (Rönnberg, 1995). The present study shows that individual and group differences do occur for visual phonetic perception.

Two sizable groups of participants were recruited, one group with normal hearing and the other with severe to profound hearing impairments. The majority of the students with hearing impairment at Gallaudet University (where part of the study was undertaken) had experienced their hearing impairments at birth or at a very early age. Hearing impairment at a young age is typically associated with lower English language and reading proficiency (Schildroth & Karchmer, 1986). Thus, if our observations showed enhanced visual speech perception among these young adults, relative to others with normal hearing, it was not likely that it was due to enhanced higher level psycholinguistic capabilities. Furthermore, evidence has been presented supporting visual perceptual specialization for manual communication, demonstrated in the deaf population having American Sign Language as a first language (e.g., Neville, 1995). Perceptual enhancement for visible speech might be predicted under similar conditions of auditory deprivation.

The present study was designed to sample lipreading across three levels of linguistic complexity in materials: consonant–vowel (CV) nonsense syllables, isolated monosyllabic words, and sentences. The experimental methods were closed-set identification of phonemes in CV nonsense syllables and open-set identification of isolated words and sentences. To our knowledge, there is not another large data base with information about visual speech perception at each of these levels across groups of adults who differed in terms of their perceptual experience (impaired vs. normal hearing). Basic descriptive adequacy seemed to us an important, yet apparently missing, basis for current investigations of speech perception involving visual as well as audiovisual conditions.

There are many methods available to study phonetic perception. The approach taken here was to obtain identification judgments and then apply various analytic tools to the obtained data. The identification of phonemes in CV syllables was considered primarily a phonetic perception task, although the task is susceptible to postperceptual response biases (see, e.g., Walden, Montgomery, Prosek, & Schwartz, 1980) and there is evidence that phoneme identification engages lexical processes (e.g., Newman, Sawusch, & Luce, 1997). The CV identification responses were scored in terms of proportion of phonemes correct and proportion of transmitted phonological feature information.

The information analyses employed sequential transmitted information (TI) analysis (SINFA, software) (Wang, 1976; Wang & Bilger, 1973) that removes the correlations among response patterns associated with features evaluated in the order of their importance. Importance is initially determined by calculating feature information nonsequentially and comparing estimates across features. Feature specifications were from Chomsky and Halle (1968; see the Appendix). In the now-classical literature on features in speech perception (see Wang & Bilger, 1973, for a review), a primary question was the psychological reality of particular feature sets. Feature analysis was not intended here to affirm the psychological reality of features, but to quantify structure in the stimulus–response

confusion matrices in terms of structural relationships among phonemes (Keating, 1988). Typically, there is not a one-to-one relationship between phonetic attributes and phonological features. But feature structure and its quantification provide evidence that phonetic cues are present in the stimuli, and this is particularly useful in evaluating responses across entire confusion matrices, particularly when responses are errorful.

Phonetic perception per se is rarely studied with word or sentence stimuli. The scoring of responses in terms of phonemes requires a methodology by which to align stimulus and response strings that can have insertions, deletions, and substitutions. For the scoring here, we employed sequence comparison software (Bernstein, Demorest, & Eberhardt, 1994) in order to align phonemically transcribed stimuli with their respective transcribed responses. The proportions of phonemes correct in isolated words and in sentences were derived from the alignment process, as was phoneme substitution uncertainty (SU), measured in bits (for a set of the sentence stimuli). The uncertainty measure employed only the phoneme error substitutions. Although lexical and other higher level processes necessarily affect the phonetic perception measures obtained with sentences, we hypothesized that error patterns could potentially provide insight into phonetic perception in sentential contexts.

METHOD

Participants

Participants with impaired hearing. Participants with impaired hearing (IH) were screened for the following characteristics: (1) an age of between 18 and 45 years; (2) enrollment at Gallaudet University; (3) sensorineural hearing impairments greater than a 60-dB HL average in the better ear across the frequencies 500, 1000, and 2000 Hz; (4) no self-report of disability other than hearing impairment, and university records reporting no impairments other than hearing impairment; (5) the self-reported use of spoken English as the primary language of the participant's family; (6) a self-report of English (including a manually coded form) as the participant's native language; (7) education in a mainstream and/or oral program for 8 or more years; and (8) vision of at least 20/30 in each eye, as determined with a standard Snellen chart.

The selection criteria were developed in order to recruit participants for whom lipreading was a socially important and well-practiced skill and to exclude participants whose native language was American Sign Language or another manual communication system other than English. Criterion (7) excluded participants who had been educated in residential schools, in which manual communication is frequently the primary mode of communication. English reading and writing abilities were estimated from the scores on the Gallaudet University English Placement Test (EPT). No normative statistics were available for this test, but it was the only a priori assessment common to every student in the population we sampled. The participants were not screened for age of onset of hearing impairment, because there was no a priori reason to do so.

Each participant completed an extensive questionnaire that addressed family and personal history of educational background, hearing impairment, and language use. Recent audiological evaluations were obtained for each participant. Students received free au-

diometric services, and if records were inadequate, the participant obtained a new evaluation.

A total of 157 individuals responded to advertisements for the experiment. One hundred twenty-four were considered potential participants, after the initial contact. Of these, 84 met the criteria outlined above, and 80 were accepted for testing. Eight IH participants were eventually excluded from the sample owing to technical problems during testing or the lack of an EPT score on file. The age range of the remaining 72 participants was 18–41 years, and 25 participants were male. The participants were paid $10 for approximately 2 h of testing.

Figure 1 is a scatterplot of the hearing levels, in better pure tone average (dB HL) across the two ears, and age of hearing impairment onset for 68 IH participants. Four participants not shown in the figure, because they became hearing impaired much later, were 7 years (115-dB HL better pure tone average), 9 years (112-dB HL better pure tone average), 17 years (140-dB HL better pure tone average), and 27 years (140-dB HL better pure tone average) of age. The majority of the participants (71%) had profound hearing impairments (90-dB HL or greater, bilaterally). Forty-five of the participants (62.5%) had hearing impairment by 6 months of age. Twenty-three (31.9%) had onsets of between 7 and 48 months. The majority experienced hearing impairment at birth (65%) or subsequently, during the period of normal language acquisition earlier than 36 months (23%). The reported causes of hearing impairment were unknown (30), meningitis (11), maternal rubella (11), other (6), genetic (5), premature birth (4), high fever (3), scarlet fever (1), and diabetic pregnancy (1). The EPT results and the questionnaire responses for these participants are reported in Bernstein et al. (1998).

Participants with normal hearing. One hundred and four participants with normal hearing (NH) were recruited at the University of Maryland, Baltimore County. They reported that hearing and vision were normal; English was identified as a first language. During the course of the experiment, data from 8 participants were excluded owing to equipment malfunction, failure to meet selection criteria, or unwillingness to complete the entire protocol. The participants ranged in age between 18 and 45 years. None of the participants reported any training in lipreading. Twenty-eight were male. They received course credit and/or $10 for their participation.

Materials

All the stimulus materials were high-quality videodisc recordings (Bernstein & Eberhardt, 1986a, 1986b) of a male and a female talker who spoke General American English. Talkers were recorded so that their faces filled the screen area. (See Demorest & Bernstein, 1992, for a description of the recordings.)

Nonsense syllables. CV *nonsense* syllables, spoken by both the male and female talkers, were composed of two tokens of each of 22 initial consonants combined with the vowel /ɑ/, plus two tokens of the vowel /ɑ/ alone. The consonants were: /p b m f v ʃ tʃ w r t θ ð d s z k g n l h dʒ ʒ/.

Words. Monosyllabic words, spoken by the male talker only, were from the clinical version (Kreul, Nixon, Kryter, Bell, & Lamb, 1968) of the Modified Rhyme Test (MRT; House, Williams, Hecker, & Kryter, 1965). The MRT comprises 50 six-word ensembles. Two words were selected from each of the ensembles, to yield 100 different words.

Sentences. Fifty sentences from the Bernstein and Eberhardt (1986a) recordings of the lists of CID Everyday Sentences (Davis & Silverman, 1970) were selected, 25 spoken by each of the talkers. An additional 25 for each talker were selected from Corpus III and Corpus IV of Bernstein and Eberhardt (1986b), a large sentence database created for the study of visual speech perception. These are referred to here as *B–E sentences*. Previous findings with these sentence materials (Demorest & Bernstein, 1992) indicated that visual perception of the female talker's speech is generally less accu-
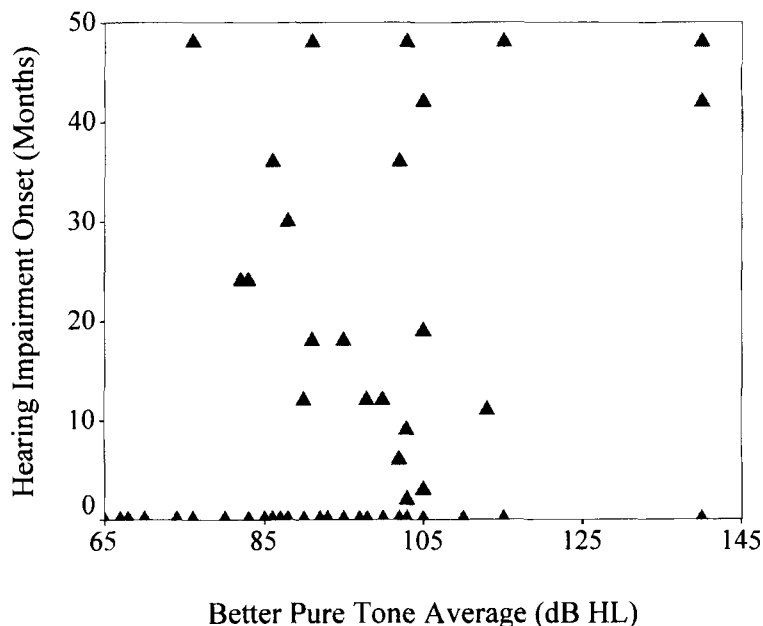
Figure 1. Distributions of the hearing levels, in better pure-tone average (dB HL) across the two ears, and age of hearing impairment onset for the 68 participants with impaired hearing.

rate than of the male talker's. In order to raise sentence scores above floor level, a set of the female talker's most intelligible sentences was selected on the basis of previous mean results. No subject saw the same sentence spoken by both talkers.

Procedures

General procedures. The availability of larger numbers of qualified NH participants afforded the possibility of studying possible practice effects within analyses of generalizability (Demorest et al., 1996). The NH participants were tested with two sets of materials, with the intertest interval ranging from 5 to 12 days. The same procedure was followed on each day, and set order was counterbalanced across participants. The materials selected for the present study made use of one set of materials only, and comparisons across IH and NH groups therefore included data from the 2nd day of testing for half of the NH participants. On the basis of the testing of the NH participants, there was no reason to suspect that results would vary significantly were both sets or only the alternate set presented.

Testing procedures were essentially the same at the University of Maryland, Baltimore County, and at Gallaudet University. The participant was seated at a small table in a darkened, sound-attenuating room. A videodisc player (Sony Lasermax LDP 1550) was controlled by a personal computer (PC). The PC was used to instruct the participant prior to testing, to present stimuli, and to record responses. A small lamp illuminated the computer keyboard. The stimuli were presented on a 19-in. high-resolution color monitor (Sony Trinitron PVM 1910), placed at a distance of 2 m from the participant. The rate of stimulus presentation was controlled by the participant, who pressed a key to see the first stimulus and pressed the return key following each subsequent stimulus and response. After a brief pause, the first frame of the stimulus was presented for 2 sec; then the remaining frames were played in real time. The final frame remained on the screen until the participant's response was completed. The monitor was darkened briefly between stimuli. The three types of materials were presented in counterbalanced order. Test sessions took from 1½ to 2 h to complete.

Nonsense syllables. The participants were tested with two 92-item lists of the CV syllables, one for each talker. Each list consisted of two repetitions of the 44 CV tokens and two repetitions of the 2 /ɑ/ tokens. Item order was pseudorandomized at presentation time, and talker order was counterbalanced across participants. Selected keys of the PC keyboard were labeled with 23 one- or two-character phonemic codes that the participant pressed to respond. The participants could not edit their responses. Prior to testing, phonemic codes were explained to the participant, and clarification was provided, if necessary. The instructions for these and the other materials were given verbally and in print for NH participants. For IH participants, verbal instructions were given in simultaneous (sign and speech) communication mode.[3] A cue card that paired each response code with an example word was visible for reference throughout the test session. The participants received a 5-item practice list.

Words. The 100 words were tested in pseudorandomized order. Prior to testing, the participants were informed that each word comprised a single syllable and that each word would be presented once. The participants were instructed to type what they thought the talker had said and were given as long as necessary to respond. Editing of the response was permitted, as well as blank responses. The participants received a five-item practice list.

Sentences. The participants viewed 25 CID sentences and 25 B–E sentences for each talker. Talker order and the order of B–E versus CID sentences were counterbalanced across NH participants.

The participants were told that they would see a series of unrelated sentences and were instructed to type exactly what they thought the talker had said. Partial responses, including word fragments, were encouraged, but the participants were instructed to keep all portions of their response in chronological order. They were also instructed to use Standard English spelling and to use contractions only when they thought that contractions were what the talker had said. The participants received a three-sentence practice set. The participants could delete and retype responses if they felt it necessary to do so. Following each response, the participants gave a con-

fidence rating on a scale ranging from 0 to 7. These ratings were collected to study in detail the relationship between subjective performance ratings and objective performance. Demorest and Bernstein (1997) showed that the two participant groups differed significantly in their subjective ratings (subjective ratings were higher among deaf participants), but an analysis of validity coefficients for the ratings showed that both groups made valid estimates of their own lipreading performance. All of those analyses included the nonresponse data. An investigation of individual differences in using subjective ratings showed that 93.0% of the validity coefficients for hearing participants and 96.4% of the validity coefficients for deaf participants were significant. Therefore, nearly all the participants made valid ratings of their own performance. These results suggest that group performance differences reported below cannot simply be attributed to response criterion differences between the two groups.

Measures

Nonsense syllables. The proportion of phonemes correct in CV syllables was the proportion of correct responses across the entire stimulus set for each of the talkers. The proportion of transmitted phonological feature information was the independent proportion of TI for the features in the Appendix, as obtained with SINFA software (Wang, 1976; Wang & Bilger, 1973).

Words. Performance on each word was scored twice. The total number of entire words correct was tallied, and a proportion of words correct was obtained. The number of phonemes correct in each word was also counted, and a proportion of phonemes correct per word was obtained. The means of these proportions were obtained for each participant.

Several steps were performed to obtain these measures. First, an experimenter checked the response files for words that were misspelled or that had obvious typographical errors (e.g., *rian* for *rain*). Corrections were made only when there was no ambiguity concerning the intended response. Because the next step was to submit response files to a computer program that counted words correct and that was sensitive to homophone distinctions, responses that contained the incorrect homophone were edited (e.g., *piece* was replaced with *peace*).

The response files were also phonemically transcribed (using a text-to-speech system) and then submitted to a sequence comparator program, which performed a phoneme-to-phoneme alignment of the stimulus with the response phoneme strings. A description and validation experiment on the sequence comparator and the overall method were described in detail in Bernstein et al. (1994). Sequence comparison techniques take into account the possibility that element-by-element alignment of two strings can require substitutions, insertions, and deletions (Sankoff & Kruskal, 1983). That is, highly similar but not identical elements can be aligned with each other, and one string may have more or fewer elements than the other. A distance metric and minimization algorithm were used to obtain the least total estimated visual perceptual distance between the stimulus and the response phoneme strings. The estimated distances between phonemes were obtained via multidimensional scaling of phoneme confusion matrices obtained in visible speech nonsense syllable identification experiments (see Bernstein et al., 1994).[4] The output of the sequence comparator included several different measures. The measure employed here was simply the number of correct phonemes per word divided by the number of phonemes in the stimulus word. Word scores were averaged across the stimulus set for each participant.

Sentences. The sentence response data were processed in a manner analogous to that for the isolated words: Files were checked for obvious spelling errors, and a computer program counted words correct per sentence. The response files were transcribed and submitted to the sequence comparator, to determine the number of phonemes correct per sentence. The phonemes correct scores per sentence

were normalized by dividing each by the number of phonemes in the stimulus sentence. In addition, the phoneme-to-phoneme alignments obtained for the CID sentences were submitted to further analysis. For example, in the following alignment:

Stimulus: p r u f r i d y u r f aɨn ḷ r ɜ s ʌ l t s
Response: b l u f – ɪ ʃ ɑ – r f ɑ n i – – – – – –

for the stimulus "proof read your final results" and the response "blue fish are funny," there are five correct phonemes and six incorrect phoneme substitutions. Extraction of incorrect substitutions was performed for each of the stimulus phonemes in the CID sentences for the male and female talkers and the two participant groups. An information measure, *SU* in bits, was calculated for each of the stimulus phonemes as

$$SU = -\Sigma p_k \log_2 p_k,$$

where p_k is the proportion of responses in category k, and k is an index of summation that represents each possible substitution error.

RESULTS AND DISCUSSION

Group and Talker Effects

An analysis of variance (ANOVA) was performed on the measures, proportion of phonemes correct in nonsense syllables, proportion of isolated words correct, mean proportion of phonemes correct in words, mean proportion of words correct in sentences, and mean proportion of phonemes correct in sentences, employing a group (IH vs. NH) × talker (male vs. female) design. The only exception to this design was that there was not a talker factor for the measures involving isolated words. Because data from the NH participants had been obtained on 2 days with counterbalanced order (see Demorest et al., 1996), preliminary analyses were conducted, and it was verified that mean performance was not significantly different for NH subgroups who viewed materials on Day 1 versus Day 2.

The means and standard deviations for each of the obtained measures for the two participant groups are given in Table 1. An ANOVA was conducted on untransformed and log-transformed scores to stabilize variances, and only for the group × talker interactions obtained with sentence stimuli were the results different. For simplicity, all the results are reported for the untransformed scores. An ANOVA on each measure showed that the group differences in the table are all significant ($p < .0005$, except for nonsense syllables, $p = .008$). The IH group consistently obtained significantly higher mean scores than did the NH group [proportion of nonsense syllables correct, $F(1,166) = 7.28$; proportion of isolated words correct, $F(1,166) = 51.33$; mean proportion of phonemes correct in isolated words, $F(1,166) = 58.34$; mean proportion of words correct in CID sentences, $F(1,166) = 62.02$; mean proportion of phonemes correct in CID sentences, $F(1,166) = 67.37$; mean proportion of words correct in B–E sentences, $F(1,166) = 71.55$; and mean proportion of phonemes correct in B–E sentences, $F(1,166) = 76.56$]. These results show that performance for the IH group was enhanced, relative to that of the NH group, for all the

Table 1
Means, Standard Deviations, and Upper Quartile Ranges for
Phoneme and Word Scores Obtained from the Impaired Hearing and Normal Hearing Groups

Group	Talker	Phoneme Scores			Word Scores		
		Mean	(SD)	Fourth Quartile Range*	Mean	(SD)	Fourth Quartile Range*
				Nonsense Syllables			
Impaired hearing	Female	.31	.06	.35–.44	NA	NA	NA
	Male	.34	.08	.39–.50	NA	NA	NA
Normal hearing	Female	.29	.07	.33–.41	NA	NA	NA
	Male	.32	.06	.37–.46	NA	NA	NA
				Isolated Words			
Impaired hearing	Male	.52	.11	.59–.73	.19	.09	.25–.42
Normal hearing	Male	.41	.09	.47–.58	.11	.06	.15–.24
				CID Sentences			
Impaired hearing	Female	.58	.20	.73–.88	.52	.20	.68–.85
	Male	.47	.18	.61–.79	.40	.18	.52–.74
Normal hearing	Female	.37	.15	.44–.69	.31	.15	.42–.75
	Male	.27	.14	.36–.64	.21	.13	.29–.49
				B–E Sentences			
Impaired hearing	Female	.43	.16	.56–.75	.35	.15	.48–.66
	Male	.53	.17	.66–.83	.47	.17	.61–.80
Normal hearing	Female	.26	.12	.27–.41	.21	.10	.26–.41
	Male	.33	.12	.42–.64	.28	.11	.36–.57

*Fourth quartile ranges do not include outliers.

types of materials and the phoneme- and word-scoring methods.

All talker differences were significant ($p < .0005$). The group $\times$ talker interactions were not significant for phoneme identification and were inconsistently significant for sentences. When the transformed scores for a particular measure produced significance, the untransformed ones did not, and vice versa. The small, although significant, interactions are not considered important. The talker difference showed the female talker to be the more difficult for the nonsense syllables and the B–E sentences. The result was reversed for the CID sentences. These results can be explained straightforwardly. As was mentioned earlier, selection of the CID sentences attempted to raise scores for the female talker and resulted in a reversal of difficulty for these materials.

Estimates of Achievable Upper Extremes on Visual Speech Perception Accuracy

Because lipreading is known to vary among individuals to an extent that has functional significance for speech communication, estimates of mean performance, although useful for group comparisons, fail to provide insights into the range of potentially functionally significant individual differences within and across populations. Of particular interest here were the upper ranges of performance: No one doubts that some individuals are poor at lipreading. The question of interest was the upper limit on visual speech perception accuracy. The upper quartiles of the response distributions were chosen to estimate achievable extremes of visual perception accuracy.[5] The upper quartiles had the advantage of including relatively large samples—approximately 18 for the IH group and 24 for

the NH group. Boxplots were found to be an excellent method for examining performance distributions and isolating the upper quartiles. Figures 2-9 show boxplots for each of the measures in the present study. In each figure, boxes represent the interquartile range, and the median is marked with a horizontal line (SPSS, 1996). Horizontal lines (at the ends of the whiskers) mark the most extreme values not considered outliers. Thus, approximately 25% of the observations fall into the range defined
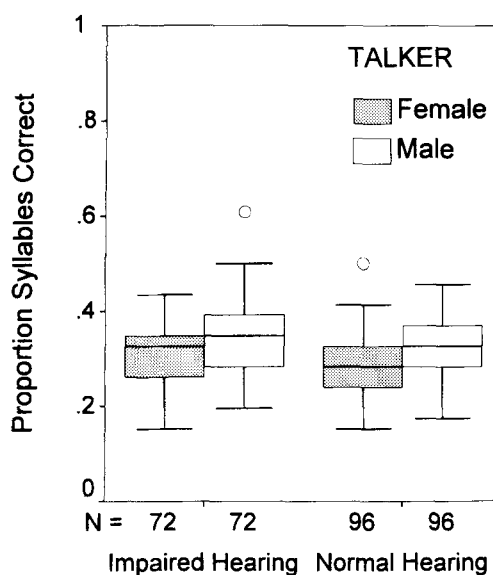

Figure 2. Boxplots of the proportion of nonsense syllables correct for impaired hearing and normal hearing participants with both the male and the female talkers.
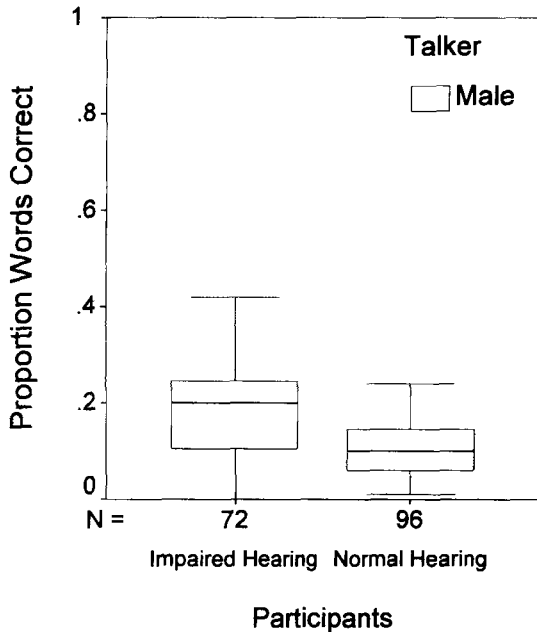
Figure 3. Boxplots of the proportion of isolated words correct for impaired hearing and normal hearing participants with the male talker.

by each whisker. Outlier scores—that is, scores more than 1.5 times the interquartile range from the ends of the box—are labeled with circles. None of the ranges discussed below includes outliers, consistent with a conservative approach to examining the upper extremes of performance.

The boxplots of proportion of nonsense syllables correct in Figure 2 show that, although there was a significant mean group difference, the interquartile range across both talkers and both participant groups covered a relatively narrow range (.24–.35). Similarly, the upper quartiles for both groups covered a narrow range (.35–.50 for the IH group and .33–.46 for the NH group), demonstrating that, even for the most proficient participants, phoneme identification in nonsense syllables is only moderately accurate. This result is consistent with the literature.

The boxplots of proportion of isolated words correct in Figure 3 show that scores were very low in both groups. It was anticipated that identification accuracy for these words would be low: Each word had at least five alternate potential response words in English, which varied either in terms of the initial consonant or consonant cluster or in terms of the final consonant or consonant cluster (see House et al., 1965). Given the high likelihood for misperception, the upper quartile performance (25%–42% correct) of the IH group can be considered remarkable. Also, the *lower* limit of the upper quartile range for the IH group coincides with the *highest* score observed for the NH group.

The boxplots of mean proportion of phonemes correct in isolated words in Figure 4 show that the upper quartile for the IH group was .59–.73 and for the NH group was .47–.58. Again, the upper quartile of the IH group

was above that of the NH group. These scores suggest a higher level of accuracy than did proportion of words correct (see Figure 3). Frequently, the participants were only partially correct in their word identifications, leading to higher scores measured in phonemes.

The boxplots of mean proportion of words correct in CID sentences are shown in Figure 5. This figure shows the discontinuity in upper quartiles for the two groups with the male talker, but not with the female talker. The upper quartile of the NH group overlaps the range of the third quartile and part of the fourth quartile of the IH group: A small subset of the NH group was as accurate as the most accurate participants in the IH group. When scored in terms of the proportion of phonemes correct, however, the upper quartile scores for the two groups are once again distinct (Figure 6).

The patterns of results are similar in Figures 7 and 8 for proportion of words correct and mean proportion of phonemes correct, respectively, for the B–E sentences. Generally, upper quartile ranges for both talkers and both scoring methods are distinct across the IH and NH groups.

All of the boxplots show that among individuals in the IH group were some who were as highly *inaccurate* at perceiving speech visually as were individuals in the NH group. Thus, severe to profound hearing impairment is not a sufficient condition for even moderately accurate visual speech perception. On the other hand, the distributions of scores, with virtually no overlap between upper quartiles for the two groups, suggest that conditions associated with auditory deprivation are favorable to enhanced visual phonetic perception.

Relationships Among Measures

Having demonstrated reliable group differences in the phoneme and word measures and individual differences
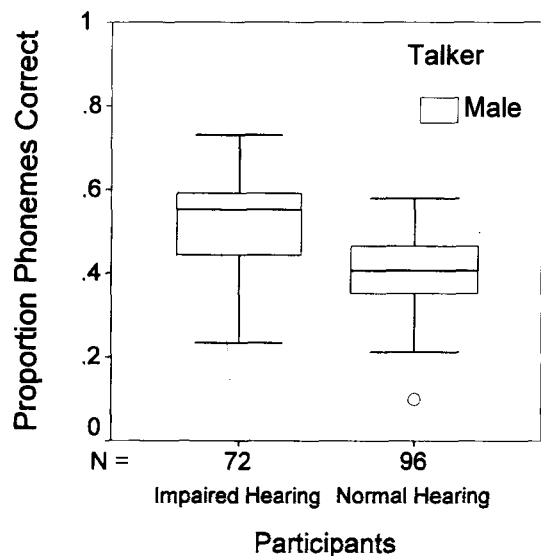

Figure 4. Boxplots of the proportion of phonemes correct in isolated words for impaired hearing and normal hearing participants with the male talker.
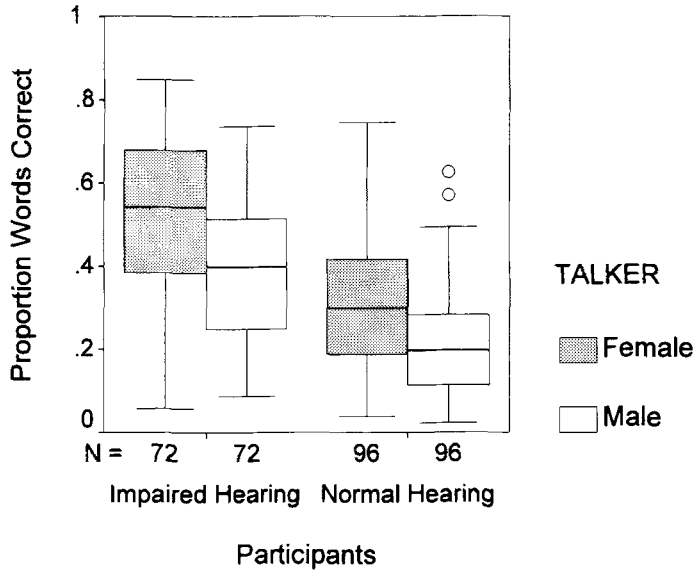
Figure 5. Boxplots of the mean proportion of words correct in CID sentences for impaired hearing and normal hearing participants with the male and female talkers.

within groups, the results were further analyzed for evidence that phonetic perception is a factor in individual differences for lipreading words and sentences. Figure 2 shows that the magnitude of the significant differences in CV identification between groups was small. Nevertheless, small increments in obtained phonetic information could result in large increments in word accuracy (Auer & Bernstein, 1997). If individuals' perceptual accuracy for phonetic information is highly implicated in individual differences in identifying isolated words and words in sentences, correlations between CV identifica-

tion and the measures involving words should be consistently substantial in magnitude.

Table 2 shows the Pearson correlations between the word scores and the CV identification scores. Isolated word scores were the proportion of words correct. Sentence scores were the mean proportion of words correct per sentence. Table 2 shows that there was a stronger association between visual phonetic perception measured with CV identification and word identification in the IH than in the NH group and that the NH participants identified the female talker's CVs in a manner unrelated to
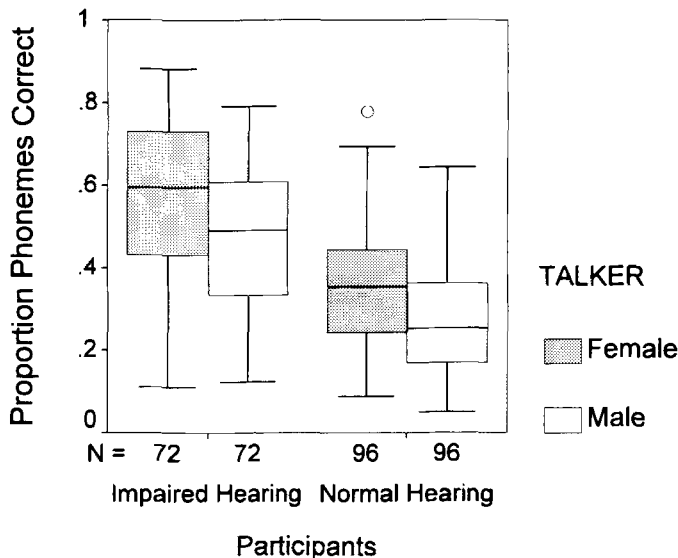

Figure 6. Boxplots of the proportion of phonemes correct in CID sentences for impaired hearing and normal hearing participants with the male and female talkers.
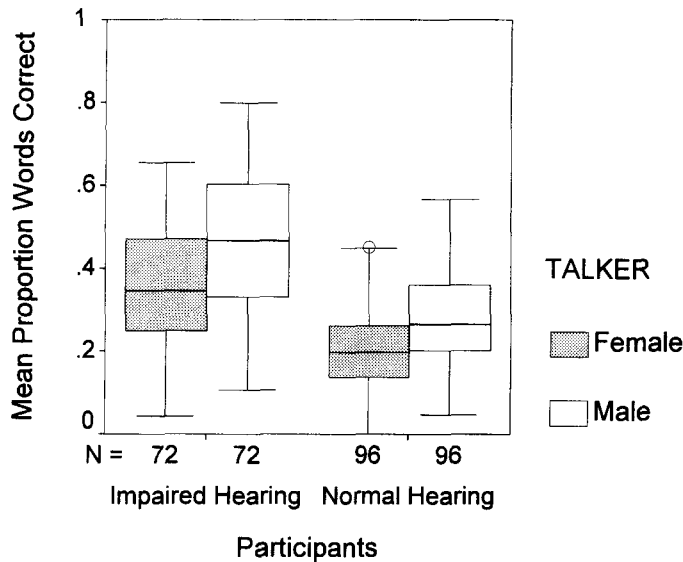
Figure 7. Boxplots of the mean proportion of words correct in B–E sentences for impaired hearing and normal hearing participants with the male and female talkers.

their identification of words. The sample correlations for the IH group were always higher than the corresponding ones for the NH group,[6] and the sample correlations with the male talker's CV identifications were always higher than those with the female talker's.

The magnitudes of the correlations of CV identification scores from the male talker and the word scores were all moderately high (.472–. 644) for the IH group. Twenty-nine percent of the variance in identifying words in the male talker's sentences is accounted for by the CV identification scores. Given that sentences afforded the pos-

sibility of guessing words, which could theoretically diminish the association between phonetic perception scores and word scores, the consistent magnitude of the correlations supports a role for phonetic perception in individual differences among the IH participants at the sentence level.

The importance of phonetic perception in accounting for individual differences is also supported when the correlations are carried out with the phoneme scoring for isolated words and sentences. Table 3 shows the Pearson correlations between the proportions of phonemes cor-
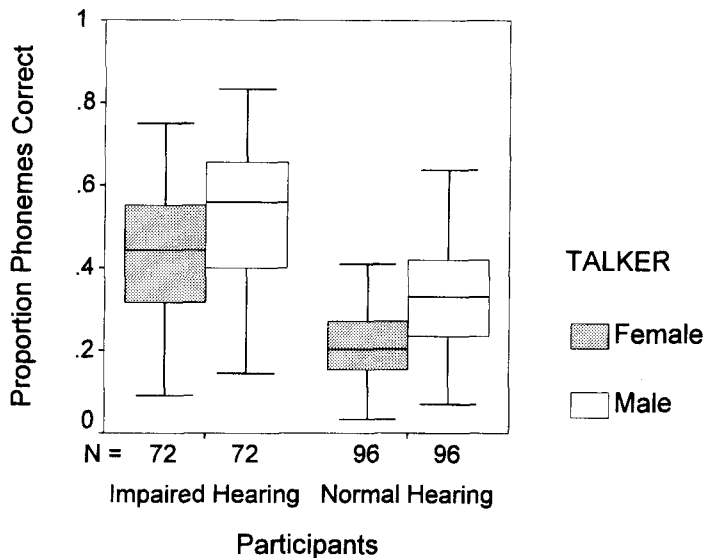

Figure 8. Boxplots of the proportion of phonemes correct in B–E sentences for impaired hearing and normal hearing participants with the male and female talkers.

Table 2
Correlations Between Proportion of Phonemes Correct in Nonsense Syllables and Proportion of Isolated Words Correct and Mean Proportion of Words Correct in Sentences for the Impaired Hearing (IH) and Normal Hearing (NH) Groups

Groups	Isolated Words, Male Talker	CID Sentences		B–E Sentences	
		Female Talker	Male Talker	Female Talker	Male Talker
	Nonsense Syllables Spoken by Female Talker				
IH	.346†	.301*	.310†	.274*	.317†
NH	.152	.111	.092	.137	.124
	Nonsense Syllables Spoken by Male Talker				
IH	.644†	.537†	.540†	.472†	.539†
NH	.389†	.426†	.430†	.378†	.404†

*p = .05.   †p = .01.

rect scores for words and sentences and the CV identification scores. Isolated word scores were the mean proportion of phonemes correct in words. Sentence scores were the mean proportion of phonemes correct in sentences. Table 3 shows that sample correlations for the IH group were invariably higher than for the NH group. The correlations for scores involving the male talker's CV identifications were higher than those involving the female talker's. The magnitude of correlations with the male talker's CV identifications was similar to those shown previously in Table 2. The similarity of the correlations across materials for a given talker and group suggests that the contribution of phonetic perception is approximately equal for both isolated word and sentence perception.

Table 3 also shows correlations between mean proportion of phonemes correct in isolated words and the phoneme scores for sentences. The proportion variance accounted for by these correlations ranged between 44% and 71% across the two participant groups. These results suggest that a large portion of the individual differences among lipreaders can be attributed to processes no later than word recognition.

The patterns of correlation between phoneme identification and phonemes correct in words and sentences is consistent with the hypothesis that variation in visual phonetic perception is a substantial portion of the explanation for individual differences in lipreading. The relatively large numbers of participants in the upper quartiles, discussed above in combination with the consistent correlational analyses, ensure that the observed high scores are not attributable merely to measurement error but, rather, to genuine occurrences of highly accurate visual speech perception.

Perception of phonological features in nonsense syllables. An analysis of the proportion of TI for phonological features (see the Appendix) was used to quantify in greater detail CV syllable identification. Feature TI was interpreted as a measure of visual phonetic cues perceived: Evidence for phonological structure implies that phonetic cues have been perceived. Feature analyses presuppose that subphonemic information is perceptible.

This presupposition is in opposition to the notion of the viseme, which is considered to be an undifferentiated perceptual unit comprising several phonemes (e.g., Massaro, 1998; Massaro, Cohen, & Gesi, 1993).

Several factors were taken into account in obtaining the TI measures. Because conditional analysis of TI by SINFA sequentially removes correlated feature structure, measures are sensitive to the order of feature evaluation. In the present study, the sparse confusion matrices (only two responses per token per talker) obtained from each participant precluded individualized analyses of feature importance. Therefore, data across groups but within talkers were pooled to obtain the relative importance (unconditional TI) of the features. Then, the conditional TI analysis was performed on data pooled only within quartiles and talkers for each participant group. Quartile assignment for each participant was based on mean proportion of phonemes correct in isolated words (see Figure 4). If phonetic perception is related to word identification, as was suggested by the correlations presented earlier, the TI measures for each of the features should be systematically related to the participants' quartile membership.

Figure 9 shows the results of the conditional proportion of TI analyses for each talker and quartile within group. Features are ordered from left to right in the figures according to the feature importance for the female talker. The first four features for the male were analyzed in the order of round, strident, coronal, and anterior but are presented here in the same order as that for the female, to facilitate direct comparison across figures. In Figure 9, the bars for each feature are ordered from first to fourth quartile, from left to right.

Figure 9 does not show the features voice, continuant, high, low, back, and duration features, which were estimated to be zero. The features in the feature set are not perfectly independent, so conditional analysis would be

Table 3
Correlations Between Proportion of Phonemes Correct in Nonsense Syllables, Mean Proportion of Phonemes Correct in Isolated Words, and Mean Proportion of Phonemes Correct in Sentences for the Impaired Hearing (IH) and Normal Hearing (NH) Groups

Groups	Isolated Words, Male Talker	CID Sentences		B–E Sentences	
		Female Talker	Male Talker	Female Talker	Male Talker
	Nonsense Syllables Spoken by Female Talker				
IH	.276*	.305†	.311†	.258*	.334†
NH	.022	.094†	.093	.032	.106
	Nonsense Syllables Spoken by Male Talker				
IH	.657†	.518†	.537†	.458†	.539†
NH	.404†	.422†	.430†	.235*	.410†
	Isolated Words Spoken by Male Talker				
IH		.836†	.843†	.812†	.843†
NH		.777†	.784†	.664†	.799†

*p = .05.   †p = .01.

A

B
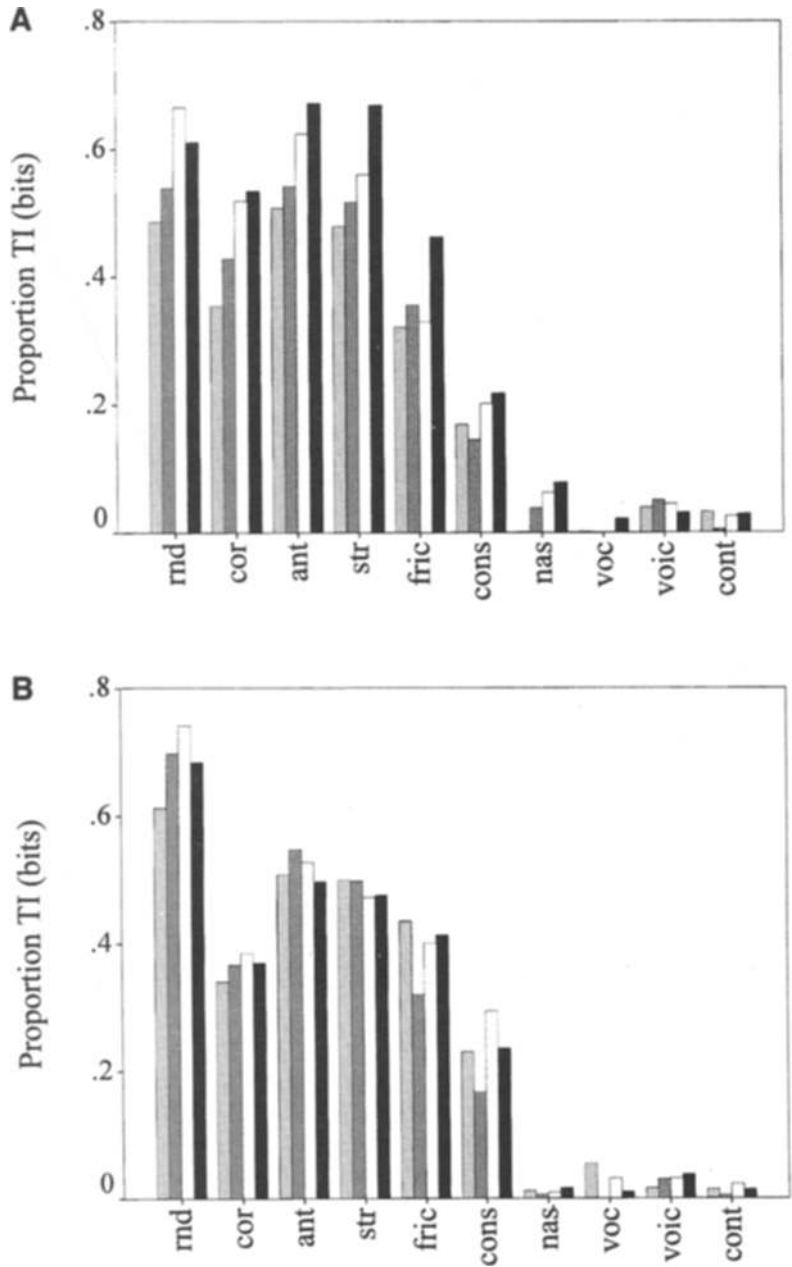
Figures 9A–9B. Proportion of transmitted information (TI) for phonological features across quartiles. Quartiles for each feature are arranged from first to fourth sequentially from left to right. Figures 9A and 9C show the impaired hearing quartiles for the female and the male talkers, respectively. Figures 9B and 9D show the normal hearing quartiles for the female and the male talker, respectively. Abbreviations: rnd, round; cor, coronal; ant, anterior; str, strident; fric, fricative; cons, consonantal; nas, nasal; voc, vocalic; voic, voicing; cont, continuant.

expected to result in a low proportion TI for some of the features. The maximum possible stimulus information for the 23 CV syllables is 4.52 bits. Were the features independent, there would be a total of 10.40 bits of stimulus information (see Wang & Bilger, 1973). Thus, this feature system contains 5.88 bits of internal redundancy. To illustrate, continuant and fricative are very similar in distribution (see the Appendix), and fricative is transmitted at a generally moderate level. On the other hand, continuant distinguishes the fricatives /ʒ ʃ/ from the affricates /dʒ ʃt/, and this distinction was not perceived. High, low, and back, with zero proportion TI, take on value only for /w a k g h ʒ ʃ dʒ ʃt/. But round is the most successfully transmitted feature and has positive value
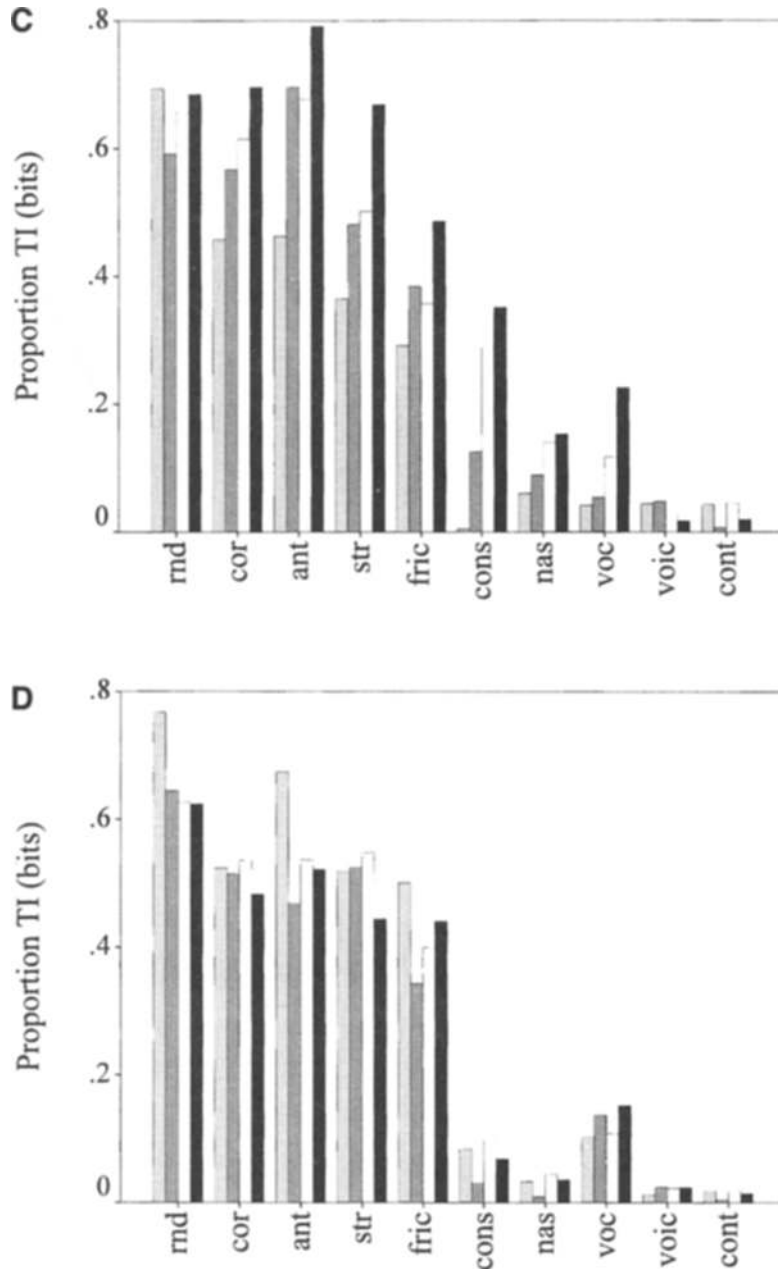
Figure 9C–9D. Proportion of transmitted information (TI) for phonological features across quartiles. Quartiles for each feature are arranged from first to fourth sequentially from left to right. Figures 9A and 9C show the impaired hearing quartiles for the female and the male talkers, respectively. Figures 9B and 9D show the normal hearing quartiles for the female and the male talker, respectively. Abbreviations: rnd, round; cor, coronal; ant, anterior; str, strident; fric, fricative; cons, consonantal; nas, nasal; voc, vocalic; voic, voicing; cont, continuant.

only for /w/. Therefore, failure to transmit a particular feature does not necessarily predict low intelligibility for particular phonemes. (See Keating, 1988, for a complete examination of phonological features.)

The main point of the TI analysis was to obtain additional evidence that performance on word identification is related to phonetic perception. The figure shows that

TI was overall higher for the IH group and for the male talker and that evidence for systematically differential TI as a function of quartile assignment was obtained only for the IH group. Figure 9A (female talker, IH group) shows a generally orderly relationship across the IH quartiles for the coronal, anterior, strident, and nasal features and somewhat less so for the round, fricative, and

consonantal features. Figure 9C (male talker, IH group) shows a wider difference between the first and the fourth IH quartiles than in Figure 9A but several points of overlap between the second and the third quartiles. Particularly large differences in proportion of TI between the first and the fourth quartiles were obtained for the coronal, anterior, strident, fricative, consonantal, and vocalic features. Nasal values are orderly with regard to quartile, but across a narrower range. Round was approximately equal across quartiles, as were voice and continuant.

Because the majority of the CV syllable identifications were errors, the TI analysis shows that more phonetic information was recovered from the stimuli by the better IH lipreaders, even when they were not absolutely correct. The contrast in pattern of results across groups suggests that individual differences in phonetic perception are more important for word identification in IH than in NH participants. This conclusion is consistent with the correlational analyses, which also showed a higher level of association between phonetic perception and identification of words in isolation and in sentences.

Phoneme substitution uncertainty in sentences. Given that phonetic perception drives word recognition, it should be possible to observe effects of phonetic perception not only in correct responses, but also in phoneme errors in sentences. Alternatively, higher level top-down processes might so strongly affect responses to sentence stimuli that phonetic perception effects might be undetectable. This issue was investigated by using the phoneme-to-phoneme alignments obtained for the CID sentences. Every instance of a stimulus phoneme aligned with a *different* phoneme (no correct responses were employed) in the response was extracted from the entire data set from the CID sentence alignments and tallied in a confusion matrix. A total of 93,455 individual phoneme-to-phoneme alignments were tallied. Matrices (four total) were compiled within groups and talkers. Phoneme *SU*, in bits, the entropy for individual phonemes, was calculated for each of the phonemes in each of the matrices.

Figure 10 shows phoneme *SU* for the female (Figure 10A) and the male (Figure 10B) talkers' CID sentences across the two participant groups. The phonetic symbols on the horizontal axes include each of the notations employed in the sentence transcriptions. Because the sentences were different across talkers, the inventory of phonetic symbols is somewhat different. The figures show that *SU*, entropy, was uniformly higher for the NH than for the IH group and higher for the female talker than for the male. An ANOVA showed that both talker and participant group were significant factors [$F(1,160) = 5.312$, $p = .022$, and $F(1,160) = 156.213$, $p = .000$, respectively].

These results show that the substitution errors committed by the NH group were less systematic than were those of the IH group. This result is predicted, if errors in the sentence responses were driven more by phonetic perception for the IH group than for the NH group. That is, we theorize that lower entropy is due to greater percep-

tual constraint. The more information obtained, the more systematic are the phoneme errors. These results are consistent with the correlational and the TI results. They extend those results by showing that group differences in phonetic perception can be detected directly in the errors in sentence responses.

GENERAL DISCUSSION

The present study supports several general conclusions. First, it shows that visual speech perception performance can be considerably more accurate (see Table 1) than commonly cited estimates, such as 30% or fewer words correct in sentences. Second, severe to profound hearing impairment can be associated with enhanced visual phonetic perception. All the evidence supports the existence of a subgroup of IH individuals who outperform adults with normal hearing. Third, relatively small differences in phonetic perception, as measured with CV syllable identification, are predictive of individual differences in isolated word and isolated sentence performance levels. Fourth, performance patterns vary across IH and NH groups. In particular, correlations among measures are higher in the IH group, differences in feature transmission scores differentiate across quartiles in the IH group but not in the NH group, and phoneme *SU* is lower in the IH group. The implications of these results are discussed below. First, however, we relate the performance levels obtained here to ones from auditory speech perception. This is to provide a more common referent (auditory) for what these performance levels might functionally accomplish. Whether performance levels equated across visual and auditory tasks of the type here accurately predict various other types of functional equivalence (e.g., ease/speed of comprehension) remains to be discovered.

Equating visual and auditory speech performance levels. The entire range of IH upper quartile scores (across all sentence sets) was .48–.85 mean proportion of words correct (see Table 1). This range can be approximately equated to the performance curve reported in G. A. Miller, Heise, and Lichten (1951) for five-word sentences presented in noise. Performance at the lower end of the range (.48) is approximately equivalent to auditory speech perception in a signal-to-noise (S/N) condition of −4 dB. Performance at the upper end of the range (.85) is approximately equivalent to performance in a +5-dB S/N condition. That is, visual speech perception in highly skilled IH lipreaders is similar in accuracy to auditory speech perception under somewhat difficult to somewhat favorable listening conditions. The entire range of NH upper quartile scores for lipreading sentences (across all the sentence sets) was .27–.69 mean proportion of words correct (see Table 1). The lower end of the range (.27) is approximately equivalent to performance in a −8-dB S/N condition. The upper end of the range (.69) is approximately equivalent to performance in a 0-dB S/N condition. That is, visual speech perception in the most
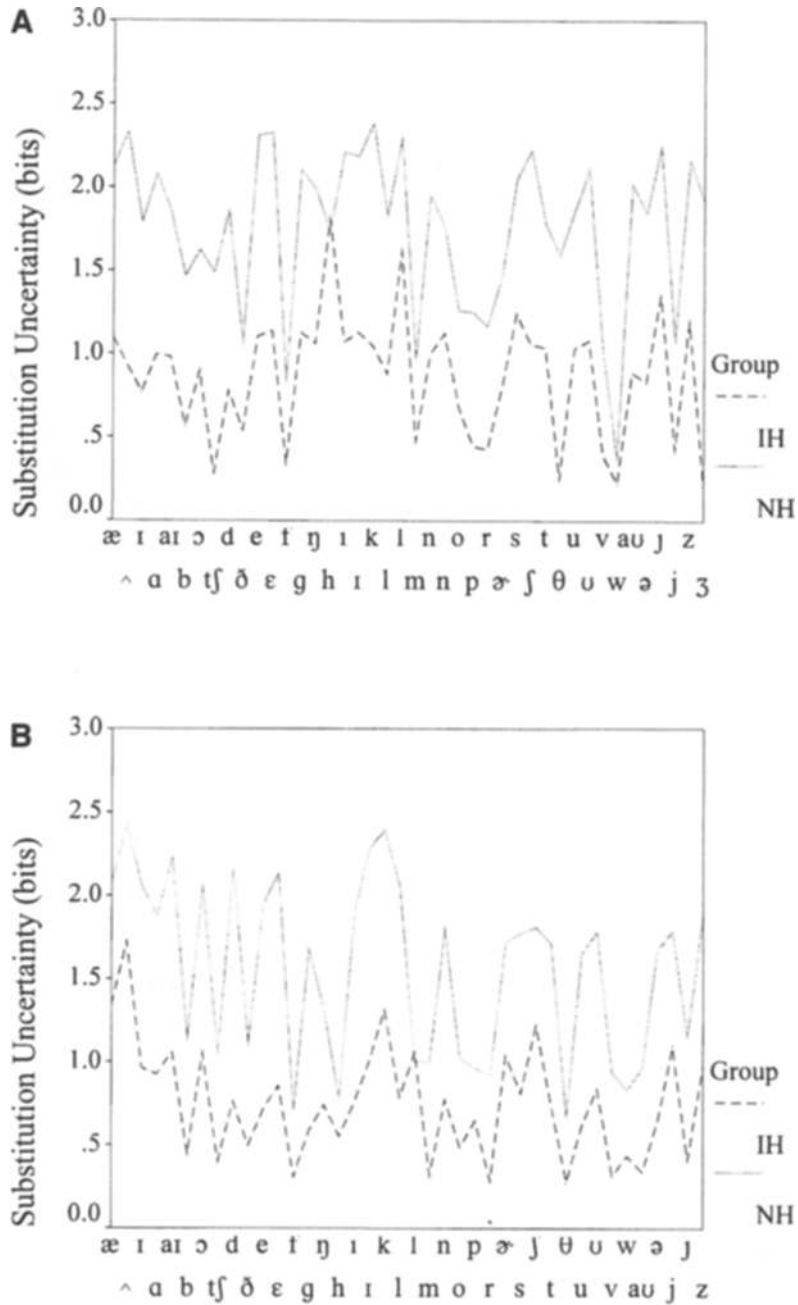
Figure 10. Substitution Uncertainty in bits for CID sentences across talkers and groups. Panel A shows the female talker, and panel B the male talker. Phonetic notation includes all the symbols required to transcribe the sentences for each of the talkers.

skilled NH lipreaders is similar in accuracy to auditory speech perception under difficult to somewhat difficult listening conditions. This characterization suggests that, for the IH upper quartile, lipreading can be adequate for fluent communication. However, for the majority of IH and NH participants, it suggests that lipreading alone renders communication difficult at best.

Impaired hearing and enhanced visual speech perception in the present study. One strong assertion in the literature is that hearing impairment is not associated with enhanced visual speech perception. For example, Summerfield (1991) states, "The distribution of lipreading skills in both the normal-hearing and hearing-impaired populations is broad. Indeed, it is to be lamented that ex-

cellence in lipreading is not related to the need to perform well, because in formal laboratory tests ... , the best totally deaf and hearing-impaired subjects often perform only as well as the best subjects with normal hearing" (p. 123). According to Rönnberg (1995), "visual speech-reading [sic] skill is not critically related to auditory impairment per se" (p. 264). (See also Mogford, 1987.) The present study provides consistent evidence for enhanced lipreading in a subset of IH participants.

In subsequent studies that we have conducted with deaf and hearing students at California State University, Northridge, at the National Center on Deafness, we have continued to find that the most accurate visual speech perception is demonstrated by the deaf students. One possible explanation for the discrepancy between our findings and the view cited above is that we recruited many participants whose severe to profound hearing impairments were acquired congenitally or very early, whereas generalizations about lipreading in IH individuals are typically based on studies of individuals with postlingual impairments and/or with less significant impairments. If enhanced visual speech perception depends on early visual perceptual experience and auditory deprivation, lipreading enhancements in postlingually deafened adults would not be predicted.

Not only is it asserted in the literature that enhanced lipreading is not associated with impaired hearing, it is also asserted that acquisition of language with normal hearing is a condition for achieving the highest levels of lipreading accuracy (Mogford, 1987). This possibility was examined with information on the present IH participants. The audiological records available for the IH group who scored in the upper quartiles on *all* of the measures were examined for evidence that hearing is *not* a prerequisite for outstanding lipreading. Four such participants had records that indicated profound, congenital impairments. Each had audiometric pure tone thresholds of 100-dB HL or greater that were probably attributable, at least in part, to vibrotactile sensation at low frequencies, and not to auditory sensation (Boothroyd & Cawkwell, 1970; Nober, 1967). Audiometric evaluations stated that these 4 participants could not identify words by listening under conditions of amplification. Absolute assurance that they had never had functionally useful hearing for identifying words is impossible. With that caveat, it appears that the upper quartile levels of lipreading accuracy were achieved by individuals with no auditory speech perception experience. The existence of these individuals suggests that visual speech perception (possibly with vibrotactile stimulation as well) can be functionally equivalent to auditory speech perception for acquiring perception of a spoken language. Future studies are needed to evaluate all aspects of spoken language processing in individuals who perform at high levels of language comprehension and production, yet have no or a minimal history of auditory speech perception. These individuals afford an opportunity to dissociate speech perception from perceptual modality in accounting for speech perception.

As was suggested above, it is likely that the present results are due to having recruited a large number of participants with severe to profound congenital or early-onset hearing impairments. If superior visual speech perception is related to severe to profound hearing impairment at an early age, prevalence of profound hearing impairment must influence the frequency of observing skilled lipreading. On the basis of 1984 statistics (Department of Commerce, 1984, cited in Hotchkiss, 1989), the percentage of the United States population 15 years of age and older who were unable to hear what is said in normal conversation was only 0.3% (481,000). Further limiting the opportunity to observe skilled lipreading is the use of American Sign Language as a first and primary language by a part of this population: Lipreading accuracy has been shown to be negatively correlated with sign language usage (Bernstein, Demorest, & Tucker, 1998; Geers & Moog, 1989; Moores & Sweet, 1991).

Phonetic Perception and Word Recognition

A principal aim of the present study was to describe visual phonetic perception across IH and NH participants. With word and sentence stimuli, it is ensured that the phonetic perception measures (phonemes correct and *SU*) are not independent of lexical and higher level psycholinguistic effects. Even CV syllable identification has been shown to be affected by lexical processes (Newman et al., 1997) and phonological knowledge (Walden et al., 1980). A question is whether the "contamination" of the phonetic perception measures by nonphonetic psycholinguistic effects diminishes the strength of the conclusion that enhanced phonetic perception occurred in the IH group. We think not.

The possibility that enhanced lipreading of sentences and words in the present study was actually attributable to enhanced higher level psycholinguistic or strategic processes might be argued on the grounds that the IH group had a better or more extensive knowledge of language to deploy during visual speech perception than did those in the NH group. However, in a recent study of similarly selected IH and NH groups, we found that the IH participants as a group are less familiar with words (Waldstein, Auer, Tucker, & Bernstein, 1996) and, in a word-age-of-acquisition task, judge words to have been learned later (Auer, Waldstein, Tucker, & Bernstein, 1996). Mean reading comprehension scores were lower, as were mean Peabody Picture Vocabulary Test (Dunn & Dunn, 1981) scores. Therefore, it seems unlikely that enhanced lipreading of sentences in the present IH group versus the NH group was due to enhanced experience or knowledge of language, although within the present IH group, there is an association between language and lipreading (Bernstein et al., 1998).

Taken together, (1) the statistically significant differences between groups on phonetic perception measures, consistently favoring the IH group, (2) the higher correlations between phonetic perception measures and word measures for the IH group, (3) the lower *SU* for phonemes

in sentences for the IH group, (4) the systematic relationships between feature transmission scores and word scores for the IH group, and (5) the high likelihood that most of the IH participants had less language experience than the NH participants all argue for genuine differences in phonetic perception across groups. Future studies, particularly ones involving measurements of optical phonetic stimulus properties and perceptual sensitivity, are needed to determine what accounts for enhanced phonetic perception.

In addition to evidence for phonetic perception differences across groups, there is strong evidence that the more successful IH lipreaders were more successful at lipreading sentences owing to their ability to recognize words. Table 3 shows correlations between mean proportion of phonemes correct in words and in sentences, which were higher for the IH group. Between 66% and 71% of the variance in sentence scores can be accounted for by the isolated word scores for the IH group, and between 44% and 64% of the variance for the NH group. Conklin (1917) and Utley (1946) both reported word versus sentence lipreading scores that correlated approximately .70. Lyxell, Rönnberg, Andersson, and Linderoth (1993) reported a similar correlation of .83.

Given that both phoneme identification and word identification were associated with sentence performance scores, we examined the predictability of CID and B–E sentence performance in terms of multiple regression analyses involving isolated word identification scores versus phoneme identification scores. Four analyses were conducted: proportions of words correct and of phonemes correct for CID sentences and for B–E sentences. The data were pooled across talkers. Potential predictors included group (IH vs. NH), phoneme identification in nonsense syllables, and isolated word identification (in the comparable metric of words correct or phonemes correct), and the interactions of group with each of the remaining predictors. The results of all four analyses were the same: Only group and isolated word performance were significant predictors of sentence performance. Phoneme identification in syllable performance did not contribute significantly to the prediction of sentence performance, once word performance was taken into account. Standardized partial regression coefficients (beta) for each two-variable regression equation (word score and group) were similar. Betas for the group variable ranged between .12 and .16. Betas for the isolated word variable ranged between .80 and .82. The multiple correlation coefficients, R, ranged between .88 and .90. These results do not mean, however, that phonetic perception is not important in explaining individual differences in lipreading sentences. They mean that the word identification measures account for the same variance as do the syllable identification scores and for additional variance not accounted for by the syllable scores.

Given the importance of word recognition in accounting for individual differences in lipreading, a question is how top-down psycholinguistic processes and knowledge versus bottom-up phonetic processing contribute to the relative accuracy of lipread word recognition across individuals. This question cannot be discussed adequately without outlining the current arguments in the literature concerning the various architectures hypothesized to account for spoken word recognition (e.g., Luce & Pisoni, 1998; Marslen-Wilson, 1990; McClelland & Elman, 1986). This exercise would far exceed the present space limitations. Even if there were space to lay out the various arguments for and against hypothesized architectures, it would need to be admitted that little work on individual differences has been done in the area of spoken word recognition (cf. Lewellen, Goldinger, Pisoni, & Greene, 1993).

In the literature on lipreading, however, the question of the relative importance of bottom-up versus top-down processing in accounting for individual differences has not been posed in terms of word recognition architectures but, rather, as a question concerning the relative importance of phonetic versus higher level psycholinguistic (sentential semantic or syntactic; Jeffers & Barley, 1971; Rönnberg, 1995; Rosen & Corcoran, 1982) or expectancy-based processing (e.g., Lyxell & Rönnberg, 1987). A similar but more extensive debate has taken place in the literature on reading (Perfetti, 1994; Stanovich, 1980) and has been raised by studies of comprehension of synthetic speech produced by rule (Duffy & Pisoni, 1992). For both modalities of language comprehension, strong evidence supports models that put a premium on bottom-up word recognition processing for fast, automatic, efficient, and accurate comprehension. The critical contact between lower level perceptual processing and stored knowledge is thought to occur during lexical processing, which makes lexical semantic and syntactic properties available to higher levels of linguistic processing (Tyler, Cobb, & Graham, 1992).

Stanovich (1980) presented the case that poorer readers are the ones who rely on semantic and syntactic expectancy-based processing, owing to their inadequate word identification skills, whereas more expert readers have fast and automatic word recognition and rely less on higher level sources of information. Stanovich suggested that "the ability to recognize words automatically is important because it frees capacity for higher-level integrative and semantic processing" (p. 60), and he argued against the notion that the more fluent readers rely more on context. Temporally prior processes of word recognition compete for resources with later comprehension processes (syntactic, semantic, discourse level). Schwantes (1981) showed that children, when they read, rely more on context than do adults, and he attributed this difference to the children's less effective word-decoding processes. Recently, Perfetti (1994) summarized the reading research on context and concluded that no studies show a greater context effect for skilled readers. Fluent reading involves automatic processes that, under normal circumstances, operate on the bottom-up stimulus. It is not the case, according to Perfetti, that fluent readers do not use context but that context aids in postidentification processes. Duffy and Pisoni (1992), share similar views in their review of perception and comprehension of speech

synthesized by rule. They argue that bottom-up phonetic perception and word recognition account for performance differences associated with different quality phonetic information. Differential use of context does not account for intelligibility differences across types of synthetic speech.

If the rule is that fluent language comprehension is driven by the accuracy, automaticity, and efficiency of word recognition, rather than by the deployment of top-down processes, it is predictable that the better lipreaders will evidence both enhanced phonetic perception and enhanced isolated word identification. Studies of lipreading are needed within the framework of detailed language comprehension processing models, with particular attention to word recognition. It would be surprising to learn that visual speech perception departs from the general form that language comprehension has been found to follow for reading and auditory speech perception.

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.

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/ESCOP Workshop on Audio-Visual Speech Processing* (pp. 21-24). Rhodes, Greece, September 26–27.

AUER, E. T., JR., WALDSTEIN, R. S., TUCKER, P. E., & BERNSTEIN, L. E. (1996). Relationships between word knowledge and visual speech perception: I. Subjective estimates of word age-of-acquisition. *Journal of the Acoustical Society of America*, **100**, 2569.

BENGUEREL, A. P., & PICHORA-FULLER, M. K. (1982). Coarticulation effects in lipreading. *Journal of Speech & Hearing Research*, **25**, 600-607.

BERNSTEIN, L. E., COULTER, D. C., O'CONNELL, M. P., EBERHARDT, S. P., & DEMOREST, M. E. (1993). Vibrotactile and haptic speech codes. In A. Risberg, S. Felicetti, G. Plant, & K.-E. Spens (Eds.), *Proceedings of the Second International Conference on Tactile Aids, Hearing Aids, and Cochlear Implants* (pp. 57-70). Stockholm, June 7-11, 1992.

BERNSTEIN, L. E., DEMOREST, M. E., COULTER, D. C., & O'CONNELL, M. P. (1991). Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America*, **90**, 2971-2984.

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: II. The psychology of speechreading and auditory–visual speech* (pp. 211-228). East Sussex, U.K.: Psychology Press.

BERNSTEIN, L. E., & EBERHARDT, S. P. (1986a). Johns Hopkins Lipreading Corpus I–II: Disc 1 [videodisc]. Baltimore: Johns Hopkins University.

BERNSTEIN, L. E., & EBERHARDT, S. P. (1986b). Johns Hopkins Lipreading Corpus III–IV: Disc 2 [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/ESCOP Workshop on Audio-Visual Speech Processing* (pp. 89-92). Rhodes, Greece, September 26–27.

BOOTHROYD, A., & CAWKWELL, S. (1970). Vibrotactile thresholds in pure tone audiometry. *Acta Otolaryngologica*, **69**, 381-387.

BREEUWER, M., & PLOMP, R. (1986). Speechreading supplemented with

auditorily presented speech parameters. *Journal of the Acoustical Society of America*, **79**, 481-499.

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

CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English*. New York: Harper & Row.

CLOUSER, R. A. (1977). Relative phoneme visibility and lipreading performance. *Volta Review*, **79**, 27-34.

CONKLIN, E. S. (1917). A method for the determination of relative skill in lip-reading. *Volta Review*, **19**, 216-219.

CONRAD, R. (1977). Lipreading by deaf and hearing children. *British Journal of Educational Psychology*, **47**, 60-65.

DAVIS, H., & SILVERMAN, S. R. (EDS.) (1970). *Hearing and deafness*. New York: Holt, Rinehart & Winston.

DEKLE, D. J., FOWLER, C. A., & FUNNELL, M. G. (1992). Audiovisual integration in perception of real words. *Perception & Psychophysics*, **51**, 355-362.

DEMOREST, M. E., & BERNSTEIN, L. E. (1992). Sources of variability in speechreading sentences: A generalizability analysis. *Journal of Speech & Hearing Research*, **35**, 876-891.

DEMOREST, M. E., & BERNSTEIN, L. E. (1997). Relationships between subjective ratings and objective measures of performance in speechreading sentences. *Journal of Speech, Language, & Hearing Research*, **40**, 900-911.

DEMOREST, M. E., BERNSTEIN, L. E., & DEHAVEN, G. P. (1996). Generalizablity of speechreading performance on nonsense syllables, words, and sentences: Subjects with normal hearing. *Journal of Speech & Hearing Research*, **39**, 697-713.

DUFFY, S. A., & PISONI, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language & Speech*, **35**, 351-389.

DUNN, L. M., & DUNN, L. M. (1981). *Peabody picture vocabulary test—revised*. Circle Pines, MN: American Guidance Service.

FEINGOLD, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist*, **50**, 5-13.

FISHER, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech & Hearing Research*, **11**, 796-804.

GEERS, A., & MOOG, J. (1989). Factors predictive of the development of literacy in profoundly hearing-impaired adolescents. *Volta Review*, **91**, 69-86.

GREEN, K. [P.] (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by eye: II. Advances in the psychology of speechreading and auditory–visual speech* (pp. 3-25). East Sussex, U.K.: Psychology Press.

GREEN, K. P., & KUHL, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics*, **45**, 34-42.

HEIDER, F., & HEIDER, G. M. (1940). An experimental investigation of lipreading. *Psychological Monographs*, **52** (Whole No. 232), 124-153.

HOTCHKISS, D. (1989). *Demographic aspects of hearing impairment: Questions and answers*. (Available from Center for Assessment and Demographic Studies, Gallaudet Research Institute, 800 Florida Avenue, N.E., Washington, DC 20002)

HOUSE, A. S., WILLIAMS, C. E., HECKER, M. H., & KRYTER, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, **37**, 158-166.

JEFFERS, J., & BARLEY, M. (1971). *Speechreading (lipreading)*. Springfield, IL: Charles C. Thomas.

KEATING, P. A. (1988). *A survey of phonological features*. Bloomington: Indiana University Linguistics Club.

KREUL, E. J., NIXON, J. C., KRYTER, K. D., BELL, D. W., & LAMB, J. S. (1968). A proposed clinical test of speech discrimination. *Journal of Speech & Hearing Research*, **11**, 536-552.

KUHL, P. K., & MELTZOFF, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy* (Vol. 20, pp. 235-266). Hillsdale, NJ: Erlbaum.

LESNER, S. A., & KRICOS, P. B. (1981). Visual vowel and diphthong perception across speakers. *Journal of the Academy of Rehabilitative Audiology*, **14**, 252-258.

LEWELLEN, M. J., GOLDINGER, S. D., PISONI, D. B., & GREENE, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, **122**, 316-330.

LISKER, L., & ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, **20**, 384-422.

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

LYXELL, B., & RÖNNBERG, J. (1987). Guessing and speechreading. *British Journal of Audiology*, **21**, 13-20.

LYXELL, B., & RÖNNBERG, J. (1989). Information-processing skill and speech-reading. *British Journal of Audiology*, **23**, 339-347.

LYXELL, B., & RÖNNBERG, J. (1991a). Visual speech processing: Word-decoding and word-discrimination related to sentence-based speech-reading and hearing-impairment. *Scandinavian Journal of Psychology*, **32**, 9-17.

LYXELL, B., & RÖNNBERG, J. (1991b). Word discrimination and chronological age related to sentence-based speechreading skill. *British Journal of Audiology*, **25**, 3-10.

LYXELL, B., & RÖNNBERG, J. (1992). The relationship between verbal ability and sentence-based speechreading. *Scandinavian Audiology*, **21**, 67-72.

LYXELL, B., RÖNNBERG, J., ANDERSSON, J., & LINDEROTH, E. (1993). Vibrotactile support: Initial effects on visual speech perception. *Scandinavian Audiology*, **22**, 179-183.

MACLEOD, A., & SUMMERFIELD, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, **24**, 29-43.

MARSLEN-WILSON, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, MA: MIT Press.

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

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

MASSARO, D. W., COHEN, M. M., & GESI, A. T. (1993). Long-term training, transfer, and retention in learning to lipread. *Perception & Psychophysics*, **53**, 549-562.

McCLELLAND, J. L., & ELMAN, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

McGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices: A new illusion. *Nature*, **264**, 746-748.

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

MILLER, G. A., HEISE, G. A., & LICHTEN, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, **41**, 329-335.

MILLER, J. L. (1991). Comment: Bimodal speech perception and the motor theory. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 139-143). Hillsdale, NJ: Erlbaum.

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.

MONTGOMERY, A. A., WALDEN, B. E., & PROSEK, R. A. (1987). Effects of consonantal context on vowel lipreading. *Journal of Speech & Hearing Research*, **30**, 50-59.

MOORES, D. F., & SWEET, C. (1991). Factors predictive of school achievement. In D. F. Moores & K. P. Meadow-Orlans (Eds.), *Educational and developmental aspects of deafness* (pp. 154-201). Washington, DC: Gallaudet University Press.

NEVILLE, H. J. (1995). Developmental specificity in neurocognitive development in humans. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 219-231). Cambridge, MA: MIT Press.

NEWMAN, R. S., SAWUSCH, J. R., & LUCE, P. A. (1997). Lexical neigh-

borhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 873-889.

NOBER, E. H. (1967). Vibrotactile sensitivity of deaf children to high intensity sound. *Laryngoscope*, **78**, 2128-2146.

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

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.

PERFETTI, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849-894). New York: Academic Press.

RABINOWITZ, W. M., EDDINGTON, D. K., DELHORNE, L. A., & CUNEO, P. A. (1992). Relations among different measures of speech reception in subjects using a cochlear implant. *Journal of the Acoustical Society of America*, **92**, 1869-1881.

RAPHAEL, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word final consonants in American English. *Journal of the Acoustical Society of America*, **51**, 1296-1303.

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

RÖNNBERG, J., ÖHNGREN, G., & NILSSON, L.-G. (1982). Hearing deficiency, speechreading and memory functions. *Scandinavian Audiology*, **11**, 261-268.

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

ROSEN, S., & CORCORAN, T. (1982). A video-recorded test of lipreading for British English. *British Journal of Audiology*, **16**, 245-254.

ROSENBLUM, L. D., & FOWLER, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 976-985.

SANKOFF, D., & KRUSKAL, J. B. (EDS.) (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Reading, MA: Addison-Wesley.

SCHEINBERG, J. C. S. (1988). *An analysis of /p/, /b/ and /m/ in the speechreading signal*. Unpublished doctoral dissertation, The City University of New York, New York.

SCHILDROTH, A. N., & KARCHMER, M. A. (1986). *Deaf children in America*. San Diego: College-Hill.

SCHWANTES, F. M. (1981). Locus of the context effect in children's word recognition. *Child Development*, **52**, 895-903.

SEKIYAMA, K., & TOHKURA, Y. (1991). McGurk effect in non-English listeners: Visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, **90**, 1797-1805.

SPSS FOR WINDOWS 7.0 [Computer Software]. (1996). Chicago, IL: SPSS Inc.

STANOVICH, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, **1**, 32-71.

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

SUMMERFIELD, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 3-52). London: Erlbaum.

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

SUMMERFIELD, Q., & McGRATH, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, **36A**, 51-74.

TILLBERG, I., RÖNNBERG, J., SVÄRD, I., & AHLNER, B. (1996). Audio-visual speechreading in a group of hearing aid users: The effects of

onset age, handicap age, and degree of hearing loss. *Scandinavian Audiology*, **25**, 267-272.

TYLER, L. K., COBB, H., & GRAHAM, N. (EDS.) (1992). *Spoken language comprehension: An experimental approach to disordered and normal processing*. Cambridge, MA: MIT Press.

UTLEY, J. (1946). A test of lip reading ability. *Journal of Speech & Hearing Disorders*, **11**, 109-116.

VATIKIOTIS-BATESON, E., MUNHALL, K. G., KASAHARA, Y., GARCIA, F., & YEHIA, H. (1996). Characterizing audiovisual information during speech. In H.T. Bunnell & W. Idsardi (Eds.), *Proceedings ICSLP 96: Fourth International Conference on Spoken Language Processing* (pp. 1485-1488). New Castle, DE: Citation Delaware.

WALDEN, B. E., ERDMAN, S. A., MONTGOMERY, A. A., SCHWARTZ, D. M., & PROSEK, R. A. (1981). Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech & Hearing Research*, **24**, 207-216.

WALDEN, B. E., MONTGOMERY, A. A., PROSEK, R. A., & SCHWARTZ, D. M. (1980). Consonant similarity judgments by normal and hearing-impaired listeners. *Journal of Speech & Hearing Research*, **23**, 162-184.

WALDEN, B. E., PROSEK, R. A., MONTGOMERY, A. A., SCHERR, C. K., & JONES, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech & Hearing Research*, **20**, 130-145.

WALDSTEIN, R. S., AUER, E. T., JR., TUCKER, P. E. & BERNSTEIN, L. E. (1996). Relationships between word knowledge and visual speech perception: II. Subjective ratings of word familiarity. *Journal of the Acoustical Society of America*, **100**, 2569.

WANG, M. D. (1976). SINFA: Multivariate uncertainty analysis for confusion matrices. *Behavior Research Methods & Instrumentation*, **8**, 471-472.

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.

WOZNIAK, V. D., & JACKSON, P. L. (1979). Visual vowel and diphthong perception from two horizontal viewing angles. *Journal of Speech & Hearing Research*, **22**, 355-365.

YEHIA, H., RUBIN, P., & VATIKIOTIS-BATESON, E. (1997). Quantitative association of orofacial and vocal-tract shapes. In C. Benoit & R. Campbell (Eds.), *Proceedings of the ESCA/ESCOP Workshop on Audio-Visual Speech Processing* (pp. 41-44). Rhodes, Greece, September 26-27.

NOTES

1. The term *deaf* is employed here to refer to individuals whose hearing impairments are at least 90-dB HL bilaterally.

2. However, vowel duration might provide a cue to postvocalic voicing distinctions (Raphael, 1972).

3. The experimenter (P.E.T.) is a certified sign language interpreter with many years of experience communicating via sign alone and via simultaneous communication.

4. It would have been possible to employ the CV identification data from the present study to derive the distance measures for the phoneme-to-phoneme sequence comparison—perhaps, by employing distance measures tuned for each of the participant groups. However, given that the goal was to compare groups, independently estimated distance measures seemed less problematic.

5. An alternative approach would have been to calculate the upper tail ratios for each of the test measures. This measure, as outlined by Feingold (1995), gives insight into the differences between groups that may arise when variances differ, independent of means. In the present study, for example, if the IH and NH groups' scores are combined into a single distribution, it can be shown that 93% of the scores in the upper 16% of the total combined population (i.e., one standard deviation above the mean) were from the IH group. The graphical approach was taken in this article because of its straightforward intuitive appeal.

6. Fisher's Z was used to statistically test the difference between correlations. Only three were significantly different. On the other hand, all sample correlations are different in the same direction across Tables 2 and 3. By the binomial test, the probability of this happening by chance is extremely small, if all population correlations are equal for both groups.

APPENDIX
Phonological Features From Chomsky and Halle (1968)

Phoneme	rnd	cor	ant	str	fric	cons	nas	voc	voi	cont	high	low	back	dur
p	0	0	1	0	0	1	0	0	0	0	0	0	0	0
t	0	1	1	0	0	1	0	0	0	0	0	0	0	0
k	0	0	0	0	0	1	0	0	0	0	1	0	1	0
b	0	0	1	0	0	1	0	0	1	0	0	0	0	0
d	0	1	1	0	0	1	0	0	1	0	0	0	0	0
g	0	0	0	0	0	1	0	0	1	0	1	0	1	0
f	0	0	1	1	1	1	0	0	0	1	0	0	0	0
θ	0	1	1	0	1	1	0	0	0	1	0	0	0	0
s	0	1	1	1	1	1	0	0	0	1	0	0	0	1
ʃ	0	1	0	1	1	1	0	0	0	1	1	0	0	1
v	0	0	1	1	1	1	0	0	1	1	0	0	0	0
ð	0	1	1	0	1	1	0	0	1	1	0	0	0	0
z	0	1	1	1	1	1	0	0	1	1	0	0	0	1
ʒ	0	1	0	1	1	1	0	0	1	1	1	0	0	1
tʃ	0	1	0	1	1	1	0	0	0	0	1	0	0	0
dʒ	0	1	0	1	1	1	0	0	1	0	1	0	0	0
m	0	0	1	0	0	1	1	0	1	0	0	0	0	0
n	0	1	1	0	0	1	1	0	1	0	0	0	0	0
r	0	1	0	0	0	1	0	1	1	1	0	0	0	0
l	0	1	1	0	0	1	0	1	1	1	0	0	0	0
w	1	0	0	0	0	0	0	0	1	1	1	0	1	0
h	0	0	0	0	1	0	0	0	0	1	0	1	0	0
a	0	0	0	0	0	0	0	1	1	1	0	1	1	1

Note—rnd, round; cor, coronal; ant, anterior; str, strident; fric, fricative; cons, consonantal; nas, nasal; voc, vocalic; voi, voicing; cont, continuant; dur, duration.