Published in final edited form as:

Ear Hear. 2001 October ; 22(5): 412–419.

Audio-Visual Perception of Sinewave Speech in an Adult Cochlear Implant User: A Case Study

Winston D. Goh, David B. Pisoni, Karen I. Kirk, and Robert E. Remez

Indiana University (W.G., D.P., K.K.), Bloomington, Indiana; and Barnard College (R.R.), New York City, New York. Winston D. Goh is now affiliated with the National University of Singapore, Department of Social Work and Psychology

Abstract

Objective—The purpose of this case study was to investigate multimodal perceptual coherence in speech perception in an exceptionally good postlingually deafened cochlear implant user. His ability to perceive sinewave replicas of spoken sentences, and the extent to which he integrated sensory information from multimodal sources was compared with a group of adult normal-hearing listeners to determine the contribution of natural auditory quality in the use of electrocochlear stimulation.

Design—The patient, “Mr. S,” transcribed sinewave sentences of natural speech under audio-only (AO), visual-only (VO), and audio-visual (A+V) conditions. His performance was compared with the data collected from 25 normal-hearing adults.

Results—Although normal-hearing participants performed better than Mr. S for AO sentences (65% versus 53% syllables correct), Mr. S was superior for VO sentences (43% versus 18%). For A+V sentences, Mr. S’s performance was comparable with the normal-hearing group (90% versus 86%). An estimate of the amount of visual enhancement, *R*, obtained from seeing the talker’s face showed that Mr. S derived a larger gain from the additional visual information than the normal-hearing controls (78% versus 59%).

Conclusions—The findings from this case study of an exceptionally good cochlear implant user suggest that he is perceiving the sinewave sentences on the basis of coherent variation from multimodal sensory inputs, and not on the basis of lipreading ability alone. Electrocochlear stimulation is evidently useful in multimodal contexts because it preserves dynamic speech-like variation, despite the absence of speech-like auditory qualities.

It has long been known that combining auditory and visual information facilitates the perception of speech. In their pioneering study of audio-visual speech perception, Sumbly and Pollack (1954) demonstrated that the intelligibility of spoken words can be enhanced by as much as +15 dB in noisy environments if listeners are able to see the talker’s face. Using visual information from a dynamically articulating face for phonetic and lexical identification benefits almost everyone when listening to speech in noisy environments, especially as people age and hearing deteriorates (Summerfield, 1987). For the hearing-impaired population, the visual modality may play a major role in everyday speech perception. Many of the speech cues for consonants that are difficult to hear are easy to see and vice versa (Walden, Prosek, Montgomery, Scherr, & Jones, 1977). For example, /f/ and /θ/ are auditorily confusable, but they are quite distinct when the talker’s articulatory

movements can be seen. The enormous gain from seeing the talker's face is eloquently captured by a question frequently asked of hearing aid practitioners—"Doctor, why can I understand you so much more clearly when I wear my glasses?" (Summerfield, 1987). Research into the nature of audio-visual speech perception may provide substantial insights and applications for rehabilitative procedures, techniques, and training methods to assist people with hearing impairment. The study of multimodal speech perception also raises many important theoretical issues about the scope and domain of current models of speech perception and spoken language processing (Bernstein, Demorest, & Tucker, 2000; Massaro, 1998; Remez, 2001).

The absolute gain in performance observed from the visual attributes of speech is highest in a noisy environment or in other conditions that make auditory perception more difficult. Accordingly, the influence of visual information has been investigated in studies of identification performance with impoverished auditory stimuli (e.g., Sumbly & Pollack, 1954). The traditional way of investigating this problem has been to manipulate the signal to noise ratio in the environment in which the speech is presented. Another way involves reducing the amount of information that is normally available in the auditory speech waveform. One such technique is to use sinewave replicas of speech instead of natural speech signals (Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981; see also Rosen, Fourcin, & Moore, 1981, for the filtered fundamental frequency technique).

In sinewave speech, time-varying sinusoidal waveforms are generated by a digital synthesizer to match the estimated center frequencies and amplitudes of the formants in a natural utterance. The synthetic sinewave pattern preserves the dynamics of frequency and amplitude variation observed in natural speech over time, but differs from natural speech in several important ways. There are no harmonics, broadband formant structures, formant frequency transitions, steady-state formants, or fundamental frequency. In short, sinewave speech patterns contain none of the "traditional" speech cues that are assumed to form the basis of speech perception, e.g., formant frequency transitions that cue manner and place of articulation (Remez et al., 1981).

Despite the unnatural characteristics of sinewave speech, these sound patterns are intelligible (Remez et al., 1981, 1994). The absence of traditional acoustic cues for phonetic perception implies that sinusoidal replicas of speech should be perceived as independently changing tones and not as integrated, linguistic perceptual objects. However, listeners are able to reliably extract the phonetic and lexical properties of the intended utterance from the highly impoverished representation of the natural token that is preserved in the sinewave replica. This phenomenon suggests that sufficient phonetic information is available in the relational and time-varying structure preserved in the sinewave pattern, even though the synthetic waveform is obviously not reproducible by a human vocal tract.

What exactly is the nature of the phonetic information encoded and represented in a sinewave replica of an utterance? In the absence of natural acoustic vocal products, the auditory qualities evoked by a sinewave sentence are highly unnatural, and no listener has ever taken an intelligible sinewave pattern as an instance of natural speech. The auditory property responsible for intelligibility is the concurrent variation of the sinusoidal components, which cohere phonetically despite unspeech-like auditory qualities because the perceiver is susceptible to the phonetically governed pattern of changes independently of the specific acoustic details that convey it.

Sinewave speech perception also shows the multimodal facilitation observed for natural speech (Remez, Fellowes, Pisoni, Goh, & Rubin, 1998). A considerable increase in

identification performance was found when the sinewave patterns were presented in an audio-visual context compared with a unimodal auditory or unimodal visual context. Multimodal facilitation of sinewave speech represents an even stronger case for phonetic coherence precisely because the presentation of a visibly articulating face with an audio track of an unnatural sounding waveform makes the event especially incongruent, as no listener would naturally attribute the talking face as the source of the sinewave sentence. Despite this, a facilitation of intelligibility with multimodal presentation over auditory alone presentation suggests an effect of sensitivity to coherent variation across the dynamical properties of the visible articulators and the audible sinewave sentence. The perceptual organization of speech therefore depends on the establishment of coherence among dissimilar elements in the auditory and visual domains (Remez et al., 1994).

The use of coherent variation in the dynamical properties of speech may be the reason that some users of cochlear implants are able to achieve exceptional levels of performance despite the distortions in the speech signal that are inherent in using the device. A “star” cochlear implant user who is also a demonstrably good lipreader might be able to use lipreading to mitigate the effects of his or her hearing impairment. However, would such a user still be able to exploit dynamic acoustic correlates of consonants and vowels in the sinewave speech? How would this user’s performance compare with normal-hearing listeners? If the “star” user is able to perform well under multimodal presentation, the findings would suggest that the cochlear implant user is perceiving the sinewave sentences on the basis of coherent variation available to normal-hearing listeners. The results would also suggest that the cochlear implant device preserves and transmits the critical time-varying dynamical pattern variation carried in sinewave sentences that appear to be sufficient for speech perception. On the other hand, if the “star” user relies mainly on his lipreading skills to recover the multimodal sinewave sentences and not on the perception of phonetic coherence, then we would expect only a modest effect of multimodal facilitation under these presentation conditions relative to the normal-hearing listeners.

In this article, we report the performance of one extraordinarily good clinical subject, Mr. S, in transcribing sinewave sentences under audio-only (AO), visual-only (VO), and audio-visual (A+V) conditions and then compare his performance with a group of normal-hearing participants whose data were collected in an earlier study by Remez et al. (1998).

Patient Background

Our patient, Mr. S, is a 35-yr-old white man with a graduate degree. He has a profound hearing loss due to cryoglobulinemia and autoimmune syndrome. Onset of his deafness occurred in 1993 when he was 29 yr old. His hearing impairment was diagnosed as a profound loss a year later and he was implanted with a Clarion 8-channel cochlear implant in 1995. He has been using the cochlear implant for the past 4 yr and is considered to be an exceptionally good user by the clinical staff at our medical center. Mr. S’s performance on a battery of standard clinical tests that were conducted while he was using the cochlear implant is described below.

The Iowa Consonant Test (Tyler, Preece, & Tye-Murray, 1983) is a closed-set test of consonant recognition in which the listener is familiarized with 16 different consonants in an /aCa/ format. The listener is then asked to identify the consonant he or she hears out of a choice of 16 alternatives. Chance performance for consonant identification is approximately 6%. This speech perception test can also be analyzed in terms of the listener’s ability to identify phonetic features. Chance performance for consonant voicing, manner, and place of articulation identification is 50%, 33%, and 20%, respectively. On the Iowa Consonant Test,

Mr. S achieved a total score of 79% correct. When scored by feature, his performance was 96% on voicing, 94% on manner, and 85% on place.

The City University of New York (CUNY) Sentences Test (Boothroyd, Hannin, & Hnath, Reference Note 1) is an open-set sentence recognition task in which the listener is presented with isolated English sentences in three listening conditions: AO, VO, and A+V. The test is scored in terms of the total number of words correctly identified. On this test, Mr. S obtained a perfect score of 100% in the A+V condition, 92% in the AO condition, and 63% in the VO condition.

Table 1 compares Mr. S's scores on these tests and the average scores of 28 other cochlear implant patients. Table 1 also lists the performance of Mr. S and the other cochlear implant patients in a recent study (Kaiser, Kirk, Pisoni, & Lachs, Reference Note 2) that measured the ability of these listeners to identify isolated consonant-vowel-consonant words from the Hoosier audiovisual multi-talker database (Lachs & Hernández, Reference Note 3; Sheffert, Lachs, & Hernandez, Reference Note 5) under AO, VO, and A+V presentations, using both single-talker and multiple-talker presentation conditions. Across all tests, Mr. S displayed substantially higher scores compared with other cochlear implant patients. Mr. S's performance on these speech perception tests shows that he is able to perceive speech without any contextual cues in a controlled test environment. It is clear from these scores that Mr. S is an exceptionally good implant user relative to the other cochlear implant patients. In particular, his lipreading performance is consistently at least 2 SDs higher than the average cochlear implant patient, as shown by his performances on the various tests in the VO conditions in Table 1.

Method

Participants

The normal-hearing participants whose data we used for comparison consisted of 25 young adults from the Indiana University community. These participants were a subset of the sample that participated in the study described in Remez et al. (1998), and were chosen because they took part in the same experimental conditions that Mr. S was tested on. The Remez et al. (1998) study included several other conditions that Mr. S was not tested on, and the participants in those conditions were not included in the comparison group. All participants were native speakers of English and reported normal hearing and normal vision or corrected-to-normal vision at the time of testing. None of the participants had previous exposure or familiarity with sinewave analogs of speech signals. All participants were enrolled in Introductory Psychology. They received either course credit for participation in this study or they were paid as a volunteer. Our patient, Mr. S, was paid as a volunteer. He also had no prior experience with sinewave speech before the present tests, and he had corrected-to-normal vision.

Apparatus and Materials

The 18 sentences used in the present study were obtained from the database developed by Remez et al. (1998) and are listed in the Appendix. The original natural sentences were recorded and digitized and then an expert phonetician analyzed the sampled data to estimate the formant center frequencies and amplitudes. Formant center frequencies were obtained by comparing discrete Fourier spectra and linear prediction estimates. The synthesis parameters were created by tracing the formant patterns over time. The fundamental frequency of phonation was estimated from a narrowband Fourier representation of the natural spectra. A software synthesizer was then used to convert the frequency and amplitude values taken at 10 msec intervals for F0, F1, F2, F3 and fricative formants to time-varying sinusoids (Rubin,

Reference Note 4). The sinewave replicas of each sentence were composed of tone analogs of the three oral formants. A fourth tone was used to reproduce fricative formants when these were present and discontinuous with the oral formants.

The first eight sentences were spoken by one of the authors (RER), and were used for the familiarization phase and AO condition. The other 10 sentences were spoken by an adult female speaker whose natural speech intelligibility had been verified by other normal-hearing volunteer participants as acoustically highly intelligible (see Bradlow, Torretta, & Pisoni, 1996). The female speaker's sentences were used in the VO and A+V conditions. The sinewave patterns for her sentences were combined and synchronized with the video samples using Adobe Premiere 4.2. All test materials were presented to participants via a Macintosh Quadra 950 machine with a Targa 2000 video card. This system presented the 14-bit color video samples at 30 frames per second in full-screen mode at 640×480 resolution on a 17-inch monitor.

Design and Procedure

The 18-sentence presentation sequence was fixed for all participants. All normal-hearing participants listened to the audio track via a pair of Beyer Dynamic DT100 headphones. The acoustic stimuli were presented at approximately 75 dB SPL. Our patient, Mr. S, listened to the audio track via a set of Labtec LS-1020 desktop computer speakers with his cochlear implant turned on. Before the start of the experiment, we calibrated the output amplitude of the speakers to a signal level (approximately 95 dB SPL) at which he could correctly identify five auditorily presented, naturally spoken consonant-vowel-consonant words in a row.

For Mr. S, the first three sentences were used as a familiarization sequence, to acclimatize him to the unnatural timbre of the sinewave sentences. These materials were presented audio-only, and the sentences were already transcribed for him on the answer sheet. Each sentence was repeated five times with 10 sec between repetitions and 20 sec between sentence blocks. A warning tone occurred before the start of a new sentence block. The next five sentences followed the same procedure and comprised the AO condition. Our patient was asked to transcribe these sentences while he listened. For the familiarization phase and the AO condition, the video monitor remained blank while the audio signals were played out via the speakers. The control participants in Remez et al. (1998) followed precisely this same procedure except that the signals were presented over individual headphones and the participants were tested in small groups of five or fewer.

In Remez et al. (1998), the VO and A+V conditions were tested between subjects using all 10 of the female talker's sentences. It was obviously not possible to follow this same procedure with our patient, and so several changes were made in the presentation format. For our cochlear implant patient, after the AO condition, the first five sentences of the female talker were presented in the VO condition, followed by the last five sentences of that same talker in the A+V condition. The cochlear implant patient and the normal-hearing control participants were instructed to look at the video monitor and to write their responses only during the intervals between repetitions and sentence blocks.

Scoring

The number of syllables correctly transcribed was used as the dependent measure for both our cochlear implant patient and the normal-hearing control group. Because there were some small procedural differences between the Remez et al. (1998) data collection and the session with our cochlear implant patient, we had to ensure that the control data obtained from the earlier study provided a valid comparison. For the control participants who were assigned to

the VO condition in Remez et al. (1998), we only scored their responses for the first five sentences, because these were exactly the same sentences presented to the cochlear implant patient in the VO condition. Conversely, for the control participants assigned to the A+V condition in Remez et al. (1998), we only scored their responses for the last five sentences, because these were the same sentences used in the cochlear implant patient's A+V condition. However, it could be argued that the control participants in the A+V condition had greater audio-visual practice because they had transcribed five sentences before the ones we scored for comparison with Mr. S. Therefore, we also scored the first five sentences for the purpose of determining whether there were significant performance differences between the two halves of the A+V controls that might suggest a practice effect. The mean performance for the first half was 80.2% (SD = 14.9) syllables correct, and was 85.6% (SD = 13.5) for the second half. A Wilcoxon matched-pairs signed-ranks test showed that this difference was not significant, $T(N = 11) = 23$, ns. Thus, there was no evidence of a practice effect for the last five sentences that we are using as a comparison for Mr. S. All five sentences were scored in the AO transcription of RER's sentences because these sentences were identical for both the cochlear implant patient and the participants in Remez et al. (1998). These scoring procedures ensured that appropriate comparisons could be made between our patient's responses and those obtained from the normal-hearing participants.

Results

The transcription performance of our cochlear implant patient, Mr. S, and the relevant data from the normal-hearing participants from Remez et al. (1998) are summarized in Table 2. Traditional parametric and nonparametric statistical testing of the differences between Mr. S and the average performance of the normal-hearing participants could not be performed because there was only a single score for the patient in each condition, and not an average of scores. As a result, there is no variance for the patient's scores. However, one way to estimate whether Mr. S was "significantly" better or worse than the normal-hearing participants is to perform a binomial test using the means of the normal-hearing controls as the test proportion. This is appropriate because the percent correct scores can be converted to dichotomous variables. A better estimate can also be obtained by using, in addition to the means, the upper and lower bounds of the 95% confidence intervals of those means as the test proportions. If the binomial tests using all three test proportions are significant, we can be confident that Mr. S's performance is outside the average range of performance of the normal-hearing controls. We report the descriptive results as well as the results of the binomial tests for each experimental condition below.

For the AO condition, the results show that Mr. S's performance (53%) was not as good as the average of the normal-hearing controls (65%). A binomial test using the controls' mean as the test proportion showed that this difference was marginally significant, $p < 0.08$. At the upper and lower bounds of the 95% confidence intervals for the control group mean, the binomial tests revealed a significant, $p < 0.01$, and nonsignificant, $p > 0.1$, difference, respectively. It should be noted that the cochlear implant patient's performance under AO presentation is, however, within 1 SD of the mean of the normal-hearing controls' performance. It is very likely that the inherent difficulty in perceiving sinewave speech is further compounded by reliance on a cochlear implant device for speech perception. Normal-hearing participants would not have as much difficulty because of their intact hearing abilities. The average performance of the normal-hearing controls displayed here is similar to the previous results reported for AO listening conditions (Remez et al., 1981).

For the VO condition, it is very clear that Mr. S showed superior performance (43%) relative to the normal-hearing controls (18%). Binomial tests using the controls' average, upper and lower bounds of the confidence intervals as the test proportions all showed a significant

difference, all $p < 0.01$. His performance was more than 2 SDs from the mean of the normal-hearing listeners. This is probably due to the increased proficiency in lipreading that is typical of the hearing-impaired population (Summerfield, 1987), although there is a lot of individual variability in lipreading skills among both normal-hearing and hearing-impaired listeners (Bernstein et al., 2000). Of greater interest to us, however, is Mr. S's performance in the A+V condition. With the addition of visual information, Mr. S's performance (90%) is comparable with the average performance of the normal-hearing controls (86%). Binomial tests using the controls' average and upper bound of the confidence interval as the test proportions showed no significant difference, $p > 0.1$, but was significant at the lower bound of the confidence interval, $p < 0.05$. Mr. S is clearly able to use and integrate information from the dynamically changing articulators in the video display with the auditory information, despite the unspeech-like qualities of the auditory sinewave speech patterns. The additional visual information improves performance to levels that are comparable with the average performance of a group of normal-hearing listeners.

One of the more interesting questions about Mr. S's performance deals with assessing the contribution of the visual information relative to the possible information available in the absence of visual stimulation (Sumby & Pollack, 1954). The actual contribution from the visual modality is the difference in scores between the A+V and AO conditions. The possible available information is the difference between the total score possible and the performance in the AO conditions, i.e., 100% minus the AO performance. The ratio of the actual contribution to the possible available information is the amount of gain obtained from seeing the talker's face, which can be considered a measure of visual enhancement. This ratio normalizes for absolute differences in AO performance and is formalized below:

$$R = (A+V - AO) / (100 - AO)$$

Binomial testing is not appropriate for the R scores because this metric is a continuous scale and does not use dichotomous variables; thus, we only report the descriptive statistics. Our patient displays a 78.3% gain from seeing the talker's face, compared with a 59.2% (SD = 28.3) gain for the mean of the normal-hearing controls. Mr. S's performance is within 1 SD of the mean of the normal-hearing controls, and suggests that he is deriving as much benefit or more from the visual information as the normal-hearing controls are in achieving his exceptionally high level of performance in the A+V condition. This level of performance is clearly not due to his superior lipreading ability alone, because he was only performing at 43% accuracy in the VO condition, although this skill is probably a contributing factor. The improvement observed under A+V presentation appears to be due to his ability to use and integrate the phonetic information provided through the visual modality with the available phonetic cues provided by the time-varying dynamics of the sinewave patterns.

Discussion

If sinewave replicas can be perceived as speech even though a natural source cannot be attributed to the signal, the results imply that the perceptual organization of speech depends critically on the establishment of coherence among dissimilar sensory patterns (Remez et al., 1994)—in the case of sinewave speech, it is the perceptual coherence of independently changing tones. In multimodal contexts, the visual perception of a dynamically articulating face combined with the auditory perception of the very unspeech-like quality of a sinewave signal makes the event especially incoherent. In other words, no natural talker can be made to appear to an observer as the original source of the sinewave speech. Therefore, multimodal presentation provides a strong test of the phonetic perceptual coherence hypothesis despite the physical dissimilarity of the visual and auditory information. In the

recent report of Remez et al. (1998), the combination of the video signal plus an auditory track containing a single tone analog of the F2 formant was the only single-tone condition that showed any facilitation over the VO condition. The interpretation of this result by Remez et al. (1998) was that both the F2 sinewave analog and the visual information together provided congruent cues regarding the underlying articulatory gestures for place of articulation, thus producing multimodal phonetic coherence of the perceptual event from two separate and seemingly independent sensory inputs.

In this study, multimodal perception of sinewave speech was investigated for the first time in an exceptionally good cochlear implant user—a “star” performer. Although Mr. S did not perform as well as the normal-hearing controls in the AO condition, his performance was comparable with the normal-hearing controls in the A+V condition. Clearly, the additional complementary visual information in the latter condition allowed both Mr. S and normal-hearing participants to perceive additional phonetic information from the sinewave sentences. Importantly, Mr. S obtained as much benefit, if not more, from seeing the talker’s face as the normal-hearing controls (78% gain versus 59% gain, respectively). This result suggests that Mr. S is able to recover the phonetic form of the underlying sentences from the coherent variation in the multimodal sensory inputs, and is not perceiving the sentences by lipreading alone.

It is also important to emphasize here that the task used in this study was to transcribe highly impoverished sinewave speech patterns, not normal or natural speech samples. Mr. S’s ability to integrate simultaneous visual information and sinewave variation in an auditory signal is impressive, and suggests that multimodal phonetic coherence can occur even when a profoundly hearing-impaired listener perceives highly reduced speech spectra through a cochlear implant device. These findings from Mr. S, a patient with a cochlear implant, provide additional support for the earlier conclusions of Remez et al. (1981)—that speech perception can endure the absence of traditional speech cues when a nonvocal acoustic carrier exhibits the time-varying properties of speech. Elements of speech perception from multiple sensory modalities depend on the establishment and preservation of perceptual coherence across the senses, and this perceptual function necessarily occurs at a level that is responsive to the common source underlying the multisensory constituents of a speech signal. Because the coherent modulations of visual and auditory samples of speech arise from the phonetically governed gestures of articulation, the perceiver’s success in integrating visual and sinewave information here arguably depends on detecting the common cause despite superficial differences in sensory quality. Moreover, the cochlear implant user is unlikely ever to place much weight, perceptually, on detecting detailed auditory qualities of speech, due to the unnatural and unspeech-like impressions of timbre evoked via electrocochlear stimulation. This kind of listener is likely to be experienced in deriving phonetic properties from the speech-like variation in a carrier of anomalous quality, as Mr. S did.

The results of this case study demonstrate the potential of research on sinewave speech perception in the hearing-impaired and other clinical populations such as the elderly and language-delayed children. The use of sinewave speech patterns with these populations may provide useful converging evidence about the similarities and differences in the nature of speech perception processes in different populations who have known and well-documented sensory, perceptual, and cognitive impairments in the ability to encode and perceive sensory input. Data from these listeners should therefore provide valuable new insights into the multimodal organization of speech and spoken language processing.

Acknowledgments

This research was supported by NIH-NIDCD Research Grant DC00111 to Indiana University, NIDCD K23 Research Grant DC00126 to Indiana University School of Medicine, NIDCD Research Grant DC00308 to Barnard College, and a staff-training grant awarded to the first author from the National University of Singapore. The authors would like to thank Stacey Yount for gathering information on the cochlear implant patients' test scores, Lorin Lachs for helpful comments during the preparation of this article, and Luis Hernandez for his invaluable technical and programming assistance.

References

- Bernstein LE, Demorest ME, Tucker PE. Speech perception without hearing. *Perception & Psychophysics*. 2000; 62:233–252. [PubMed: 10723205]
- Bradlow AB, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine grained acoustic-phonetic talker characteristics. *Speech Communication*. 1996; 20:255–272. [PubMed: 21461127]
- Massaro, DW. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press; 1998.
- Remez, RE. The interchange of phonology and perception seen from the perspective of organization. In: Hume, EV.; Johnson, KA., editors. *The Role of Speech Perception Phenomena in Phonology*. New York: Academic Press; 2001. p. 27-52.
- Remez RE, Fellowes JM, Pisoni DB, Goh WD, Rubin PE. Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication*. 1998; 26:65–73. [PubMed: 21423823]
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychological Review*. 1994; 101:129–156. [PubMed: 8121955]
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981; 212:947–950. [PubMed: 7233191]
- Rosen SM, Fourcin AJ, Moore BCJ. Voice pitch as an aid to lipreading. *Nature*. 1981; 291:150–152. [PubMed: 7231534]
- Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*. 1954; 26:212–215.
- Summerfield, Q. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd, B.; Campbell, R., editors. *Hearing By Eye: The Psychology of Lip-Reading*. Hillsdale, NJ: Erlbaum; 1987. p. 3-51.
- Tyler, RS.; Preece, J.; Tye-Murray, N. *The Iowa Cochlear Implant Tests*. Iowa City, IA: Department of Otolaryngology—Head and Neck Surgery, University of Iowa; 1983.
- Walden BE, Prosek RA, Montgomery AA, Scherr CK, Jones CJ. Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*. 1977; 20:130–145. [PubMed: 846196]

Reference Notes

1. Boothroyd, A.; Hannin, L.; Hnath, T. A sentence test of speech perception: Reliability set equivalence and short-term learning (internal report RCI 10). New York: City University of New York; 1985.
2. Kaiser, A.; Kirk, KI.; Pisoni, DB.; Lachs, L. Audiovisual speech integration in adults with cochlear implants or normal hearing: Lexical and talker effects. Paper presented at the ARO midwinter meeting; St. Petersburg Beach, F. L. February 20–24; 2000.
3. Lachs, L.; Hernández, LR. Update: The Hoosier audiovisual multi-talker database (Research on Spoken Language Processing Progress Rep No 22. Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory; 1998. p. 377-388.
4. Rubin, PE. Internal memorandum. New Haven, CT: Haskins Laboratories; 1980. Sinewave synthesis.

5. Sheffert, S.; Lachs, L.; Hernandez, LR. The Hoosier audiovisual multi-talker database (Research on Spoken Language Processing Progress Rep No 21. Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory; 1997. p. 578-583.

Appendix

Sentence Materials (adapted from Remez et al., 1998)

Familiarization

The bill was paid every third week.

The soft cushion broke the man's fall.

Two blue fish swam in the tank.

Audio-only (AO)

A small creek cut across the field.

The fruit peel was cut in six slices.

Her purse was filled with useless trash.

The stray cat gave birth to kittens.

Where were you a year ago?

Visual-only (VO)

Always close the barn door tight.

This is a grand season for hikes on the road.

He ran halfway to the hardware store.

Kick the ball straight and follow through.

The term ended in late June that year.

Audio-visual (A+V)

Use a pencil to write the first draft.

Cut the pie into large parts.

The boy was there when the sun rose.

A cup of sugar makes sweet fudge.

What joy there is in living.

TABLE 1

Comparison of Mr. S's percentage scores and the average percentage scores of other cochlear implant patients on several speech perception tests.

	Other Cochlear Implant Patients (N = 28)		
	Mr. S	M	SD
Iowa Consonant Test ^a			
Total	79	45.0	17.3
Voicing	96	88.7	12.9
Manner	94	67.1	15.7
Place	85	52.3	16.3
CUNY Sentences Test ^b			
Audio-only	92	55.0	29.6
Visual-only	63	24.3	15.6
Audio-visual	100	91.2	9.1
Indiana Multi-talker Isolated CVC Words ^c			
Single talker condition			
Audio-only	67.0	30.1	19.1
Visual-only	30.5	15.7	5.1
Audio-visual	88.5	69.0	13.8
Multiple talker condition			
Audio-only	55.5	30.0	16.6
Visual-only	30.5	14.8	7.1
Audio-visual	86.0	60.0	14.6

^aTyler, Preece, and Tye-Murray (1983).

^bBoothroyd, Hannin, and Hnath (1985).

^cData from Kaiser, Kirk, Pisoni, and Lachs (2000).

TABLE 2

Average percentage of correct syllables for Mr. S and normal-hearing listeners for the three presentation conditions.

	Audio-Only	Visual-Only	Audio-Visual
Mr. S	52.5	43.2	89.7
Normal-hearing ^a			
M	64.7	18.0	85.6
95% confidence intervals	54.6–74.8	11.9–24.2	76.5–94.6
SE	4.9	2.8	4.1
SD	24.4	10.6	13.5
<i>N</i>	25	14	11

^aFrom Remez, Fellowes, Pisoni, Goh, and Rubin (1998).